DOCUMENT RESUME

ED 418 095		TM 028 189

AUTHOR	Ashcraft, Alyce S.
TITLE	Ways To Evaluate the Assumption of Multivariate Normality.
PUB DATE	1998-04-11
NOTE	67p.; Paper presented at the Annual Meeting of the Southwestern Psychological Association (New Orleans, LA, April 11, 1998).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Chi Square; *Evaluation Methods; *Multivariate Analysis
IDENTIFIERS	Graphic Representation; *Normality Tests; *Scattergrams

ABSTRACT

This paper reviews graphical and nongraphical methods for estimating multivariate normality. Prior to exploring this methodology, a foundation is established by presenting ways to assess univariate and bivariate normality. A data set of three variables used by J. Stevens (1986) is analyzed using Q-Q plots, stem and leaf plots, histograms, skewness, and kurtosis coefficients, the Shapiro-Wilk statistic, and bivariate and multivariate scatterplots. Multivariate normality is explored in terms of calculating Mahalanobis distances and plotting them on a scattergram against derived chi-square values using Fortran and Statistical Package for the Social Sciences (SPSS) programs developed by B. Thompson (1990, 1997). Appendixes, which comprise more than half the half, contain the SPSS commands, two computer programs for the analysis, and some results of the analyses. (Contains 24 figures and 11 references.) (Author/SLD)

Running head: MULTIVARIATE NORMALITY

Ways to Evaluate the Assumption of Multivariate Normality

Alyce S. Ashcraft

The University of Texas at Austin

Paper presented at the annual meeting of the Southwestern Psychological Association,

New Orleans, April 11, 1998

Abstract

The present paper reviews the graphical and nongraphical methods for estimating multivariate normality. Prior to exploring this methodology, a foundation will first be established by presenting ways to assess univariate and bivariate normality. A data set of three variables used by Stevens (1986) is analyzed using Q-Q plots, stem and leaf plots, histograms, skewness and kurtosis coefficients, the Shapiro-Wilk statistic, and bivariate and multivariate scatterplots. Multivariate normality is explored in terms of calculating Mahalanobis distances and plotting them on a scattergram against derived chi-square values using Fortran and SPSS programs developed by Thompson (1990, 1997).

Ways to Evaluate the Assumption of Multivariate Normality

Multivariate analyses are vital to the social sciences in the exploration of a dynamic environment. Fish (1988) and Thompson (1994) stated that use of multivariate methods are vital for two reasons. First, multivariate methods avoid the inflation of experimentwise Type I error rates that occur when univariate methods are employed in a single study to test multiple hypotheses that are at least partially uncorrelated. Secondly, and more importantly, multivariate methods analytically honor a substantive reality in which most effects have multiple causes and multiple consequences.

The trend toward utilization of multivariate methods has increased over the past two decades, as noted by Emmons, Stallings, and Layne (1990) and Grimm and Yarnold (1995). The former group of researchers studied 16 years of research reports in three journals and found that

> the multivariate characteristic of the social science research environment with its many confounding or intervening variables has been addressed through the trend toward increased use of multivariate analysis of variance and covariance, multiple regression, and multiple correlation. (p. 14)

The latter group of researchers noted that, "In the last 20 years, the use of multivariate statistics has become commonplace. Indeed, it is difficult to find empirically based articles that do not use one or another multivariate analysis" (p. vii).

Because these methods are gaining in popularity, it is important to understand the assumptions underlying multivariate statistical techniques, one of which is multivariate normality. It is imperative to remember that multivariate normality is basic to the statistical significance inference procedure of multivariate analysis (Marascuilo& Levin, 1983). The purpose of the

present paper is to review the graphical and nongraphical methods for estimating multivariate normality. Prior to exploring this methodology, a foundation will first be established by presenting ways to assess univariate and bivariate normality.

Normality

Parametric tests require the estimation of a least one population parameter from the sample statistics. To make the estimation, certain assumptions must be made, the most important of which is that the variable measured in the sample is *normally* distributed in the population to which it is to be generalized (Munro & Page, 1993). It is important to remember that the normal curve is a mathematical model that depends upon the mean and the standard deviation, in the restrictive sense that the mean and the standard deviation are used to calculate skewness and kurtosis. Skewness and kurtosis quantitatively evaluate the normality of the distribution, with skewness referring to the asymmetry of the curve and kurtosis referring to the tallness or flatness of the curve (Bump, 1991).

Properties of the Normal Curve. The properties of the normal curve include the following:

1. The curve is symmetrical. The mean, median, and mode coincide.

2. The maximum ordinate of the curve occurs at the mean, that is, where z = 0 in a normal z score distribution, and the unit normal curve is equal to .3989.

3. The curve is asymptotic. It approaches but does not meet the horizontal axis and extends from minus infinity to plus infinity.

4. The points of inflection of the curve occur at points plus or minus one standard deviation unit above and below the mean. Thus the curve changes from convex to concave in relation to the horizonal axis at these points.

5. Roughly 68% of the area of the curve falls within the limits plus or minus one standard deviation unit from the mean.

6. In the unit normal curve the limits z = +/- 1.96 include 95% and the limits z = +/- 2.58 include 99% of the total area of the curve, 5% and 1% of the area, respectively, falling beyond these limits.

(Ferguson, 1976, p. 98)

Univariate Normality

Before proceeding to a discussion of multivariate normality, it is important to review univariate and bivariate normality because "normality on each of the variable is a necessary but not sufficient condition for multivariate normality to hold" (Stevens, 1996, p. 243). Analysis of variance (ANOVA) tests whether between group means differ and has as one of its assumptions that the dependent variable should be normally distributed. ANOVA is robust with respect to the normality assumption and skewness has very little effect (generally only a few hundredths) on level of significance or power if the design is "balanced" (i.e., equal number of observations per cell). Platykurtosis (flattened distribution relative to the normal distribution) attenuates power (Stevens, 1996).

Univariate tests for assessing normality may be graphical and nongraphical. To graphically determine univariate normality, a Q-Q Plot (quantile-versus-quantile), compares observed values

with expected normal distribution values. In these plots, scores are ranked and sorted. An expected normal value is computed and compared with the actual normal values for each case. The expected normal value is the position a case with that rank holds in a normal distribution; the normal value is the position it holds in the actual distribution. If the actual distribution is normal, the points for the cases fall along the diagonal running from lower left to upper right, with some minor deviations secondary to random processes (Tabachnick & Fidell, 1989).

Figure 1 graphically displays a variable with one hundred responses in increasing order of magnitude plotted against expected normal distribution values. Normality is tenable in this instance because the plot resembles a straight line. Figure 2 is an arrangement of 50 responses for a variable in increasing order of magnitude plotted against expected normal distribution values. Normality is not tenable in this instance because the plot does not resemble a straight line. Only two points are plotted when $n = 50$. In this instance, other pictorial representations assist in the determination of normality.

Q-Q plots are available using the graphs menu on SPSS (Appendix A). SPSS also provides stem and leaf plots (e.g., Figure 3) and histograms (e. g., Figure 4) for visualization of normality. The normal curve, as presented in basic statistical texts, is more readily visualized in stem and leaf plots and histograms. Figures 3 and 4 demonstrate the classic bell curve using the one hundred responses denoted in Figure 1. Figures 5 and 6 fail to demonstrate normality using the 50 responses denoted in figure 2. It is important to remember that with small or moderate sample sizes, it may be difficult to tell whether graphic non-normality is real or apparent (Gnanadesikan, 1977; Neter, Kutner, Nachtsheim, & Wasserman, 1996; Norusis, 1995).

The most powerful non-graphic tests for determining univariate normality includes the

skewness and kurtosis coefficients and the Shapiro-Wilk test (Stevens, 1996). In SPSS, this information can be obtained with the Explore procedure (Appendix A). Note that SPSS will print the Shapiro-Wilk for samples with less than 50 observations and the K-S Lilliefors statistic for samples with greater than 50 observations. Table 1 shows the SPSS Descriptives printout for data with 100 responses and Table 2 shows the SPSS Descriptives printout for data with 26 responses.

Fisher's Measure of Skewness. This statistic is based on deviations from the mean to the third power. A symmetrical curve will result in a value of 0. If the skewness value is positive, then the curve is skewed to the right, and vice versa. Dividing the measure of skewness by the standard error for skewness results in a number that is interpreted in terms of the normal curve. Values above +1.96 or below -1.96 are statistically significant because 95% of the scores in the normal distribution fall between +1.96 and -1.96 standard deviations from the mean. Because this statistic is based on deviations to the third power, it is very sensitive to extreme values (Munro & Page, 1993). The coefficients in Tables 1 and 2 are not statistically significant.

Fisher's Measure of Kurtosis. This statistic indicates whether a distribution is too flat or too peaked, being based on deviations of the mean to the fourth power. If the kurtosis value is positive, the distribution is too peaked to be normal; if the kurtosis value is negative, the curve is too flat to be normal. The kurtosis statistic is divided by the standard error for kurtosis and the values compared to the +/- 1.96 range used to determine skewness (Munro & Page, 1993). The coefficients in Tables 1 and 2 are not statistically significant.

Shapiro-Wilk Test. Shapiro and Wilk developed a test for normality that is sensitive to a wide variety of alternatives to the normal. Small values of W correspond to departure from

normality. If observed significance levels are reasonably large (greater than 0.1), normality is not an unreasonable assumption (Gnanadesikan, 1977). The Shapiro-Wilk statistic in Table 2 is sufficiently large so that the assumption of normality is tenable.

Bivariate Normality

The normal correlation model for the case of two variables is based on the bivariate normal distribution. Consider the vocabulary (X_1) scores and math (X_2) scores for a group of students from Table 3. The student's score combinations form a scatter diagram (Figure 7). The centroid, ($X_1 = 17.6$. $X_2 = 16.1$), is the center of the 10 cases (Tatsuoka, 1971b). If there was a large population of students, a clustering of points would be expected around the centroid with a gradual thinning as the distance away from the centroid continues. To depict this in a manner analogous to the normal curve, a third dimension, frequency, is needed perpendicular to the (X_1, X_2) plane.

The surface will resemble a bell shaped "mound" similar to Figures 8, 9, 10, and 11, with the apex vertically above the centroid (Karson, 1982, Neter, Kutner, Nachtsheim, & Wasserman, 1996, Tatasuoka, 1971a, 1971b). For every pair of values (X_1, X_2), the density $f(X_1, X_2)$ represents the height of the surface at that very point. The surface is continuous, with probability corresponding to the volume under the surface (Neter, Kutner, Nachtsheim, & Wasserman, 1996). Though this conveys a general impression, it is customary to represent the bivariate curve with a series of contour lines. These contour lines (Figure 12) are a series of concentric ellipses and their common center is the centroid. The statistical implication of the volume under the bivariate normal surface of a given elliptical region is parallel to the meaning of the area under the normal curve over a given interval. It represents the probability that a random bivariate

observation, when plotted as a point on the (X_1, X_2) plane, will lie within the elliptical region. For example, in Figure 12, an observation that falls in the small ellipse has an 80% chance of being included in the sample because it is close to the mean, whereas an observation that falls in the large ellipse has a 20% chance of being included in the sample because it is far from the mean (Morrison, 1983). The contour is a cross section of the surface made by a plane parallel to the (X_1, X_2) plane. Thinking must still be three dimensional because the bell shaped "mound" is being sliced into sections, with the top part of the "mound" being the top of the normal curve and the bottom part of the "mound" being the bottom of the normal curve. Thus, bivariate normality is checked by graphing X_1 and X_2 and noting the scatter of the variables around the centroid. The pattern should be elliptical (Karson, 1982, Neter, Kutner, Nachtsheim, & Wasserman, 1996, Tatasuoka, 1971a, 1971b).

Multivariate Normality

Multivariate normality is assessed to verify the reasonableness of assuming normality for a given body of multiresponse questions. As can be imagined, there are many possibilities for departure from normality with multiresponse data. A preliminary step in evaluating the normality of multiresponse data is to evaluate univariate normality for each of the variables. In the printout of the MULTINOR Program written by Thompson (1990) (Appendix B), univariate normality for each of the three variables was checked using Q-Q Plots, stem and leaf plots, histograms, the Shapiro-Wilk's statistic, and skewness and kurtosis coefficients (Figures 13 through 21; Tables 4 and 5). The Q-Q plots of the three variables (Figures 13, 14, and 15) show that normality is tenable for variable one because the plot resembles a straight line but normality is not as tenable for variables two and three because the plots do not resemble a straight line. The stem and leaf

plot and histogram of variable one (Figures 16 and Figure 19) reveal a somewhat normal distribution while the stem and leaf plots and histograms of variables two (Figures 17 and 20) and three (Figures 18 and 21) reveal negatively skewed and trimodal distributions respectively. The descriptives data (Tables 4 and 5) reveal skewness and kurtosis statistics that are not statistically significant for all three variables and Shapiro-Wilk statistics that are significantly large for variables one and three to make the normality assumption not unreasonable. Univariate normality cannot be assumed for these variables. Remember that univariate normality was discussed because "normality on each of the variables separately is a necessary, *but not sufficient*, condition for multivariate normality to hold" (Stevens, 1996, p. 243).

Next, for normality to hold, any linear combinations of the variables must be normally distributed and all subsets of the set of variables must have multivariate normal distributions. This condition implies that all pairs of variables must be bivariate normal (Stevens, 1996). Bivariate normality was checked for in the MULTINOR data (Appendix B) by requesting scatterplots and noting elliptical patterns for the three possible combinations of the variables (Figures 22 through 24). A cursory view of the patterns around the centroids does not reveal a clear elliptical pattern. Measuring and connecting the variables to form elliptical patterns based on percentages (80%, 60%, 40%, and 20%) of variables around the centroid assists in visualizing the ellipses.

The data can finally be checked for multivariate normality by calculating the Mahalanobis distance (D^2) for each subject (Thompson, 1990). The Mahalanobis distance is the distance of a case from the centroid of the remaining cases where the centroid is the point defined by the means of all the variables (Tabachnick & Fidell, 1989). Basically, it indicates how far a case is from the centroid of all cases for the predictor variables. A large distance indicates an observation that is

an outlier for the predictors. The Mahalanobis distance is the accepted measure of distance between two (quantitative) multivariate populations and is independent of sample size (Krzanowski, 1988; Stevens, 1996).

In the MULTINOR printout, (Appendix B) the D^2 can be calculated for each subject using the formula $D^2_i = (x_i - x)' \; S^{-1} \; (x_i - x)$ where x_i is the vector of data for case i and x is the vector of means (centroid) for the predictors. Using the data for subject eight from the MULTINOR printout, the equation for subject eight would be as follows (numbers are rounded to the nearest tenth):

$$D^2_8 = (.3, -0.9, 0.5) \begin{pmatrix} 0.57 & -0.12 & -0.37 \\ -0.12 & 0.33 & -0.26 \\ -0.37 & -0.26 & 0.92 \end{pmatrix} \begin{pmatrix} 0.3 \\ -0.9 \\ 0.5 \end{pmatrix} = 0.69408$$

1 x 3 3 x 3 3 x 1 1 x 1

Based on the formula, the matrices are 1 x 3, 3 x 3, and 3 x 1. To determine the numbers for the equation, first subtract the mean of each variable from the scores of the selected subject to form the 1 x 3 and 3 x 1 matrices and use the inverted variance/covariance matrix from the printout for S^{-1}. The results will match the Mahalanobis distances given on the second page of the MULTINOR printout. After the distances are calculated, the values are sorted in ascending order and paired with a derived chi-square value [(j - 0.5)/n = percentile for the chi-square]. A table or computer program is required to determine p values because each chi square is not at the standard 0.01 or 0.05 levels (see the second page of the MULTINOR printout). The pairs are then plotted in a scattergram (see the third page of the MULTINOR printout). If n (number of subjects in the sample) - p (number of variables) is greater than 25, the plot should resemble a straight line.

Conceptually, it is important to remember that the inverted variance/covariance matrix serves as a constant in the equation. Just by looking at the 1 x 3 and 3 x 1 matrices and their relation to the centroid, deciding where a subject will fall on a graph is possible. Order inferred distance can be estimated without the inverted variance/covariance matrix.

Looking at the MULTINOR scatterplot (Appendix B), each subject can be identified. Subject 8 is the first * in the lower left hand corner because the D^2/chi square value is the closest to the centroid; subject 17 is the * in the far upper right hand corner because the D^2/chi square value is fartherest from the centroid (0/0). Again, distance indicates how far the case is from the centroid and if the plot resembles a straight line, normality is more tenable. The Mahalanobis distance represents the coordinate for the three means. In a multivariate normal curve, the cases will cluster around the centroid and taper off as the distance increases.

Thompson (1997) wrote an SPSS program to test multivariate normality graphically (Appendix C). Note the commands on the first page of the program. Page two of the program lists all of the variables for the data set and their means. On page three of the program, the Mahalanobis statistics are listed with the residual statistics. Page four details the Mahalanobis Distances for each subject in ascending order (subject number six is first; subject number three is last). The distances are paired with Chi Square values and graphed (page six).

Homogeneity of Variance-Covariance Matrices

An indirect way to assess multivariate normality is to test the assumption that the variance-covariance matrices within each cell of the design are sampled from the same population variance-covariance matrix. If the matrices are sampled from the same population, they can reasonably be pooled to create a single estimate of error. Evaluation of homogeneity of variance-

covariance matrices in especially important when sample sizes are not equal.

SPSS MANOVA conducts a Box's M test to determine homogeneity of the variance-covariance matrices. The null hypothesis for the Box's M test is that the variance-covariance matrices are not statistically significant, therefore a p value of greater that 0.05 is desired. If the assumption for multivariate generalization of homogeneity of variance is met, then it is likely that the assumption for multivariate normality is also met. This paper will not discuss in depth the relationship between normality and homogeneity and refers the reader to Tabachnick and Fidell (1989) for further exploration.

Conclusion

Although multivariate normality is not required to estimate most multivariate parameters (e.g., function coefficients, structure coefficients), even in these cases the distributions of the variables must be reasonably comparable. To test for multivariate normality, univariate and bivariate assumptions should be met in addition to calculating Mahalanobis distances and plotting them against a derived chi-square value to note their linearity. If the assumption for multivariate normality is met solely through calculation of Mahalanobis distances and graphically noting linearity, then the assumptions for univariate and bivariate normality are met. However, if data are determined to be univariate and bivariate normal, it may not be assumed to be multivariate normal. Computer programs are available to ease calculations to determine normality, including Thompson's Multinor (1990, 1997) program.

References

Bump, W. (1991, January). The normal curve takes many forms: A review of skewness and kurtosis. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX. (ERIC Document Reproduction Service No. ED 342 790)

Emmons, N.J., Stallings, W.M., & Layne, B.H. (1990, April). Statistical methods used in *American Educational Research Journal, Journal of Educational Psychology,* and *Sociology of Education* from 1972 through 1987. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 319 797).

Ferguson, G. (1976). Statistical analysis in psychology and education (4th ed.). New York: McGraw-Hill Book Company.

Fish, L. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Grimm, L.G., & Yarnold, P.R. (Eds.). (1995). Reading and understanding multivariate statistics. Washington, D.C.: American Psychological Association.

Gnanadesikan, R. (1977). Methods for statistical data analysis of multivariate observations. New York: John Wiley & Sons.

Karson, M. (1982). Multivariate statistical methods: An introduction. Ames, IA: The Iowa State University Press.

Krzanowski, W. (1988). Principles of multivariate analysis. New York: Oxford University Press.

Marascuilo, L. A., & Levin, J. R. (1983). Multivariate statistics in the social sciences: A researcher's guide. Monterey, CA: Brooks/Cole.

Morrison, D. (1983). Applied linear statistical methods. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Munro, B., & Page, E. (1993). Statistical methods for health care research (2nd ed.). Philadelphia: J. B. Lippincott Company.

Neter, J., Kunter, M., Nachtsheim, C., & Wasserman, W. (1996). Applied linear statistical models (4th ed.). Chicago: Irwin.

Norusis, M. (1995). SPSS: SPSS 6.1 Guide to data analysis. Englewood Cliffs, NJ: Prentice Hall.

Stevens, J. (1986). Applied multivariate statistics for the social sciences (2nd ed.). Mawah, NJ: Lawrence Erlbaum Associates.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mawah, NJ: Lawrence Erlbaum Associates.

Tabachnick, B., & Fidel, L. (1989). Using multivariate statistics (2nd ed.). New York: HarperCollins Publishers.

Tatsuoka, M. (1971a). Multivariate analysis: Techniques for educational and psychological research. New York: John Wiley and Sons.

Tatsuoka, M. (1971b). Selected topics in advanced statistics: An elementary approach. Champaign, IL: The Institute for Personality and Ability Testing.

Thompson, B. (1990). MULTINOR: A Fortran program that assists in evaluating multivariate normality. Educational and Psychological Measurement, 50, 845-848.

Thompson, B. (1994, February). Why multivariate methods are usually vital in research: Some basic concepts. Paper presented as a featured speaker at the biennial meeting of the Southwestern Society for Research in Human Development, Austin, TX. (ERIC Document Reproduction Service No. ED 367 687)

Thompson B. (1997). MULTINOR program for SPSS. [Computer programming language]. College Station, TX: Author.

Appendix A

SPSS Commands

```
PPLOT
 /VARIABLES=one
 /NOLOG
 /NOSTANDARDIZE
 /TYPE=Q-Q
 /FRACTION=BLOM
 /TIES=MEAN
 /DIST=NORMAL.
GRAPH
 /HISTOGRAM=one.
EXAMINE
 VARIABLES=one two three
 /PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
 /COMPARE GROUP
 /STATISTICS DESCRIPTIVES
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.


GRAPH
 /SCATTERPLOT(BIVAR)=one WITH three
 /MISSING=LISTWISE.
PLOT
 /VERTICAL='VARIABLE ONE' REFERENCE (6.4)
 /HORIZONTAL='VARIABLE THREE' REFERENCE (6.7)
 /PLOT=ONE WITH THREE.
GRAPH
 /SCATTERPLOT(BIVAR)=one WITH two
 /MISSING=LISTWISE.
PLOT
 /VERTICAL='VARIABLE ONE' REFERENCE (6.4)
 /HORIZONTAL='VARIABLE TWO' REFERENCE (6.9)
 /PLOT=ONE WITH TWO.
GRAPH
 /SCATTERPLOT(BIVAR)=two WITH three
 /MISSING=LISTWISE.
PLOT
 /VERTICAL='VARIABLE TWO' REFERENCE (6.9)
 /HORIZONTAL='VARIABLE THREE' REFERENCE (6.7)
 /PLOT=TWO WITH THREE.
```

Appendix B

```
******PROGRAM MULTINOR,                                             VERS 1.0
      WRITTEN BY BRUCE THOMPSON

      LOGIC FOR METHOD FROM STEVENS, 1986, PP. 207-212

JOB TITLE: Analysis of Stevens Example p. 209

DECLARED NUMBER OF GROUPS IS:        2

N OF SUBJECTS IN GROUP 1 IS:        26

N OF SUBJECTS IN GROUP 2 IS:        12

TOTAL N OF SUBJECTS IS:             38

N OF VARIABLES IS:                   3

INPUT DATA FILE IS:                 99

FORMAT WITH WHICH DATA TO BE READ: (3F5.1)

DATA MATRIX FOR GROUP #1

     1     5.80000     9.70000     8.90000
     2    10.60000    10.90000    11.00000
     3     8.60000     7.20000     8.70000
     4     4.80000     4.60000     6.20000
     5     8.30000    10.60000     7.80000
     6     4.60000     3.30000     4.70000
     7     4.80000     3.70000     6.40000
     8     6.70000     6.00000     7.20000
     9     7.10000     8.40000     8.40000
    10     6.20000     3.00000     4.30000
    11     4.20000     5.30000     4.20000
    12     6.90000     9.70000     7.20000
    13     5.60000     4.10000     4.30000
    14     4.80000     3.80000     5.30000
    15     2.90000     3.70000     4.20000
    16     6.10000     7.10000     8.10000
    17    12.50000    11.20000     8.90000
    18     5.20000     9.30000     6.20000
    19     5.70000    10.30000     5.50000
    20     6.00000     5.70000     5.40000
    21     5.20000     7.70000     6.90000
    22     7.20000     5.80000     6.70000
    23     8.10000     7.10000     8.10000
    24     3.30000     3.00000     4.90000
    25     7.60000     7.70000     6.20000
    26     7.70000     9.70000     8.90000

VARIABLE MEANS:
           6.40385     6.86923     6.71538   Centroid

VAR/COV MATRIX:

     1     4.52279     3.98212     2.94114
     2     3.98212     7.41261     3.70049
```

3	2.94114	3.70049	3.31015

INVERTED VAR/COV MATRIX:

1	0.56740	-0.12024	-0.36973
2	-0.12024	0.33075	-0.26292
3	-0.36973	-0.26292	0.92454

MAHALANOBIS DISTANCES WITHIN GROUP #1

1	5.40434
2	5.89352
3	2.67146
4	1.30686
5	2.38155
6	1.79599
7	2.75163
8	0.69408
9	1.19431
10	4.90097
11	2.41353
12	1.77029
13	2.80861
14	1.28071
15	2.75819
16	2.00243
17	10.53041
18	3.92584
19	7.68053
20	0.82928
21	1.40629
22	0.94312
23	1.42369
24	2.71666
25	1.72773
26	1.78797

SORTED D SQ AND ASSOCIATED CHI-SQUARE VALUES AND p
WITH DF=3 AND PERCENTILE = 100(I - .5)/N

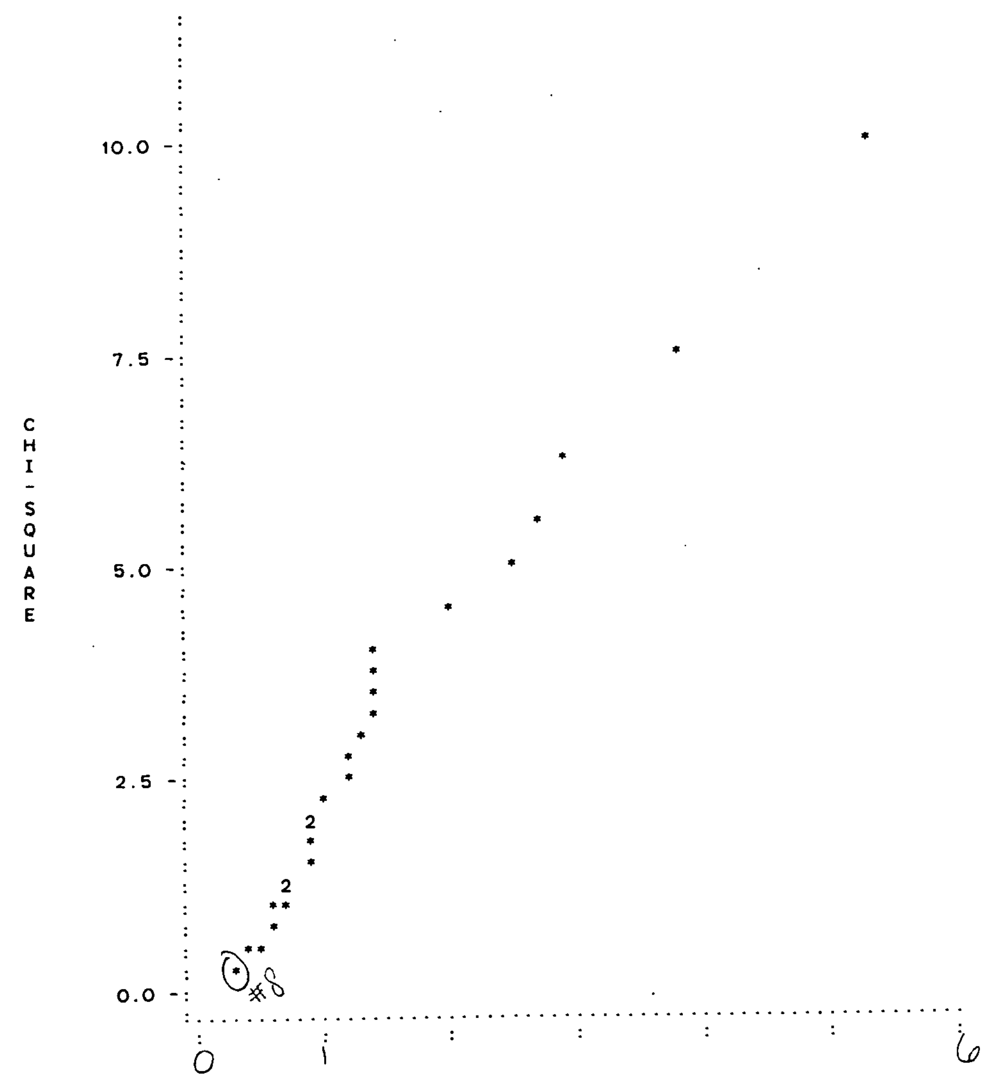
	D Sq	chi sq	p
1	0.69408	0.17988	0.01923 —#8
2	0.82928	0.38996	0.05769
3	0.94312	0.56743	0.09615
4	1.19431	0.73313	0.13462
5	1.28071	0.89380	0.17308
6	1.30686	1.05287	0.21154
7	1.40629	1.21253	0.25000
8	1.42369	1.37444	0.28846
9	1.72773	1.53997	0.32692
10	1.77029	1.71044	0.36538
11	1.78797	1.88716	0.40385
12	1.79599	2.07154	0.44231
13	2.00243	2.26515	0.48077
14	2.38155	2.46983	0.51923
15	2.41353	2.68779	0.55769
16	2.67146	2.92176	0.59615
17	2.71666	3.17526	0.63462
18	2.75163	3.45290	0.67308
19	2.75819	3.76095	0.71154
20	2.80861	4.10835	0.75000

BEST COPY AVAILABLE

21	3.92584	4.50845	0.78846
22	4.90097	4.98259	0.82692
23	5.40434	5.56822	0.86538
24	5.89352	6.34088	0.90385
25	7.68053	7.49482	0.94231
26	10.53041	9.92311	0.98077

SCATTERPLOT OF D SQ AND CHI SQUARE FOR GROUP #1

1



BEST COPY AVAILABLE

0. 2. 4. 6. 8. 10. 12.

MAHALANOBIS DISTANCE

DATA MATRIX FOR GROUP #2

1	2.40000	2.10000	2.40000
2	3.50000	1.80000	3.90000
3	6.70000	3.60000	5.90000
4	5.30000	3.30000	6.10000
5	5.20000	4.10000	6.40000
6	3.20000	2.70000	4.00000
7	4.50000	4.90000	5.70000
8	3.90000	4.70000	4.70000
9	4.00000	3.60000	2.90000
10	5.70000	5.50000	6.20000
11	2.40000	2.90000	3.20000
12	2.70000	2.60000	4.10000

VARIABLE MEANS:

4.12500	3.48333	4.62500

VAR/COV MATRIX:

1	1.91659	0.96318	1.64386
2	0.96318	1.31606	1.05864
3	1.64386	1.05864	1.97659

INVERTED VAR/COV MATRIX:

1	1.85081	-0.20446	-1.42975
2	-0.20446	1.35758	-0.55706
3	-1.42975	-0.55706	1.99335

MAHALANOBIS DISTANCES WITHIN GROUP #2

1	2.59346
2	2.53196
3	5.85428
4	2.37118
5	1.98854
6	0.70038
7	2.22173
8	2.17303
9	5.59246
10	3.12622
11	1.65042
12	2.19634

SORTED D SQ AND ASSOCIATED CHI-SQUARE VALUES AND p
WITH DF=3 AND PERCENTILE = 100(I - .5)/N

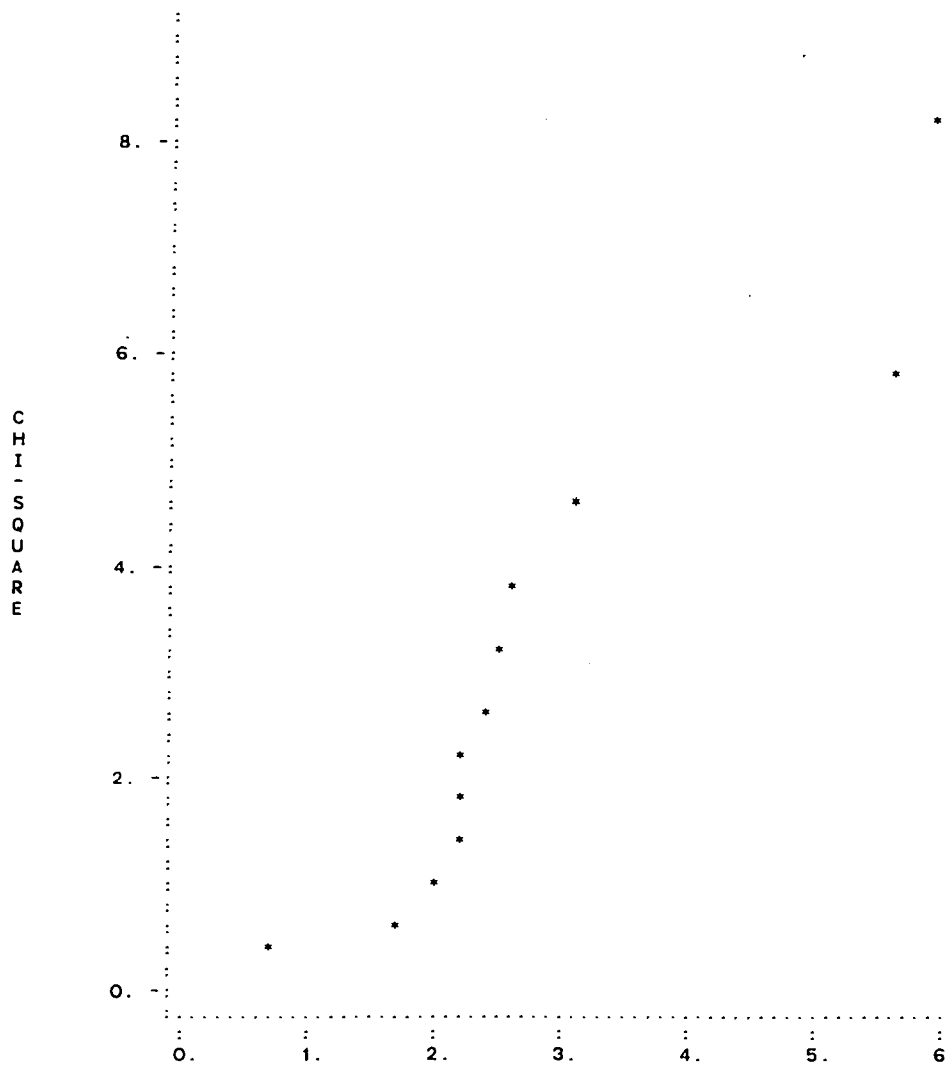
	D Sq	chi sq	p
1	0.70038	0.30897	0.04167
2	1.65042	0.69236	0.12500
3	1.98854	1.03962	0.20833
4	2.17303	1.38807	0.29167
5	2.19634	1.75398	0.37500
6	2.22173	2.15099	0.45833
7	2.37118	2.59519	0.54167

BEST COPY AVAILABLE

8	2.53196	3.10983	0.62500
9	2.59346	3.73392	0.70833
10	3.12622	4.54475	0.79167
11	5.59246	5.73942	0.87500
12	5.85428	8.22058	0.95833

SCATTERPLOT OF D SQ AND CHI SQUARE FOR GROUP #2

1




MAHALANOBIS DISTANCE

Appendix C

multino2.aer 10/11/97

multinor.sps

```
SET BLANKS=SYSMIS UNDEFINED=WARN printback=list.
TITLE 'MULTINOR.SPS   tests multivar normality graphically****'.
COMMENT *******************************************************.
COMMENT The original MULTINOR computer program was presented,
COMMENT with examples, in:
COMMENT     Thompson, B. (1990). MULTINOR: A FORTRAN program that
COMMENT     assists in evaluating multivariate normality.
COMMENT     _Educational and Psychological Measurement_, 50,
COMMENT     845-848.
COMMENT
COMMENT The logic and the data source for the example are from:
COMMENT     Stevens, J. (1986). _Applied multivariate statistics
COMMENT     for the social sciences. Hillsdale, NJ: Erlbaum.
COMMENT     (pp. 207-212)
COMMENT ***********************************************************.
COMMENT  Here there are 3 variables for which multivariate
COMMENT  normality is being confirmed.
DATA LIST
  FILE='c:\spsswin\multinor.dat' FIXED RECORDS=1 TABLE
  /1 x1 1-3 (1) x2 5-7 (1) x3 9-11 (1).
list variables=all/cases=9999/format=numbered .
COMMENT  'y' is a variable automatically created by the program, and
COMMENT  does not have to modified for different data sets.
compute y=$casenum .
print formats y(F5) .
regression variables=y x1 to x3/
  descriptive=mean stddev corr/
  dependent=y/enter x1 to x3/
  save=mahal(mahal) .
sort cases by mahal(a) .
execute .
list variables=x1 to x3 mahal/cases=9999/format=numbered .
COMMENT  In the next TWO lines, for a given data set put the actual
COMMENT  in place of the number '12' used for the example data set.
loop #i=1 to 12 .
COMMENT  In the next line, change '3' to whatever is the number
COMMENT  of variables.
COMMENT      The p critical value of chi square for a given case
COMMENT  is set as [the case number (after sorting) - .5] / the
COMMENT  sample size].
compute p=($casenum - .5) / 12. .
compute chisq=idf.chisq(p,3) .
end loop .
print formats p chisq (F8.5) .
list variables=y p mahal chisq/cases=9999/format=numbered .
plot
  vertical='chi square'/
  horizontal='Mahalabis distance'/
  plot=chisq with mahal .
```

multinor.dat

```
2.4 2.1 2.4
3.5 1.8 3.9
6.7 3.6 5.9
5.3 3.3 6.1
5.2 4.1 6.4
3.2 2.7 4.0
4.5 4.9 5.7
3.9 4.7 4.7
4.0 3.6 2.9
5.7 5.5 6.2
2.4 2.9 3.2
2.7 2.6 4.1
```

```
multinor.lst

-> SET BLANKS=SYSMIS UNDEFINED=WARN printback=list.

-> TITLE 'MULTINOR.SPS   tests multivar normality graphically****'.

-> COMMENT ***********************************************************.
-> COMMENT The original MULTINOR computer program was presented,
-> COMMENT with examples, in:
-> COMMENT     Thompson, B. (1990). MULTINOR: A FORTRAN program that
-> COMMENT     assists in evaluating multivariate normality.
-> COMMENT     _Educational and Psychological Measurement_, 50,
-> COMMENT     845-848.
-> COMMENT
-> COMMENT The logic and the data source for the example are from:
-> COMMENT     Stevens, J. (1986). _Applied multivariate statistics
-> COMMENT     for the social sciences. Hillsdale, NJ: Erlbaum.
-> COMMENT     (pp. 207-212)
-> COMMENT ****************************************************************.

-> COMMENT  Here there are 3 variables for which multivariate
-> COMMENT  normality is being confirmed.

-> DATA LIST
->   FILE='c:\spsswin\multinor.dat' FIXED RECORDS=1 TABLE
->   /1 x1 1-3 (1) x2 5-7 (1) x3 9-11 (1).

-> list variables=all/cases=9999/format=numbered .

        X1   X2   X3
    1  2.4  2.1  2.4
    2  3.5  1.8  3.9
    3  6.7  3.6  5.9
    4  5.3  3.3  6.1
    5  5.2  4.1  6.4
    6  3.2  2.7  4.0
    7  4.5  4.9  5.7
    8  3.9  4.7  4.7
    9  4.0  3.6  2.9
   10  5.7  5.5  6.2
   11  2.4  2.9  3.2
   12  2.7  2.6  4.1
```

Variables

```
Number of cases read:  12    Number of cases listed:  12

-> COMMENT  'y' is a variable automatically created by the program, and
-> COMMENT  does not have to modified for different data sets.

-> compute y=$casenum .

-> print formats y(F5) .

-> regression variables=y x1 to x3/
->   descriptive=mean stddev corr/
->   dependent=y/enter x1 to x3/
->   save=mahal(mahal) .

           * * * *   M U L T I P L E   R E G R E S S I O N   * * * *
Listwise Deletion of Missing Data
           Mean  Std Dev  Label
Y          6.500    3.606
X1         4.125    1.384
X2         3.483    1.147
X3         4.625    1.406
```

N of Cases = 12

Correlation:

	Y	X1	X2	X3
Y	1.000	-.207	.376	-.044
X1	-.207	1.000	.606	.845
X2	.376	.606	1.000	.656
X3	-.044	.845	.656	1.000

* * * * M U L T I P L E   R E G R E S S I O N * * * *

Equation Number 1   Dependent Variable.. Y
Descriptive Statistics are printed on Page 83

Block Number 1. Method: Enter X1 X2 X3

Variable(s) Entered on Step Number
1.. X3
2.. X2
3.. X1

Multiple R	.66417
R Square	.44112
Adjusted R Square	.23154
Standard Error	3.16069

44% of the observed variability is "explained" by the independent variables

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	3	63.08053	21.02684
Residual	8	79.91947	9.98993

F = 2.10480   Signif F = .1780

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
X1	-1.909097	1.296480	-.733029	-1.473	.1791
X2	2.445453	1.110369	.778083	2.202	.0588
X3	.165296	1.345478	.064454	.123	.9053
(Constant)	5.092203	3.454771		1.474	.1787

End Block Number 1 All requested variables entered.

* * * * M U L T I P L E   R E G R E S S I O N * * * *

Equation Number 1   Dependent Variable.. Y

Residuals Statistics:

	Min	Max	Mean	Std Dev	N
*PRED	2.0801	9.9172	6.5000	2.3947	12
*ZPRED	-1.8457	1.4270	.0000	1.0000	12
*SEPRED	1.2118	2.4798	1.7932	.3534	12
*ADJPRED	.6074	10.6661	6.2406	2.9511	12
*RESID	-5.0425	5.0265	.0000	2.6954	12
*ZRESID	-1.5954	1.5903	.0000	.8528	12
*SRESID	-1.9334	1.8781	.0291	1.0420	12
*DRESID	-7.4057	7.0104	.2594	4.0901	12
*SDRESID	-2.4778	2.3496	.0287	1.2152	12
*MAHAL	.7004	5.8543	2.7500	1.5070	12
*COOK D	.0000	.4543	.1364	.1713	12
*LEVER	.0637	.5322	.2500	.1370	12

Total Cases = 12

```
          * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
From Equation    1:    1 new variables have been created.

  Name        Contents
  ----        --------
  MAHAL        Mahalanobis' Distance

-> sort cases by mahal(a) .
-> execute .

-> list variables=x1 to x3 mahal/cases=9999/format=numbered .

        X1   X2   X3        MAHAL
   1   3.2  2.7  4.0       .70038  Sub 6
   2   2.4  2.9  3.2      1.65042
   3   5.2  4.1  6.4      1.98854
   4   3.9  4.7  4.7      2.17303
   5   2.7  2.6  4.1      2.19634
   6   4.5  4.9  5.7      2.22174
   7   5.3  3.3  6.1      2.37118
   8   3.5  1.8  3.9      2.53196
   9   2.4  2.1  2.4      2.59346
  10   5.7  5.5  6.2      3.12622
  11   4.0  3.6  2.9      5.59246
  12   6.7  3.6  5.9      5.85428  Sub 3

Number of cases read:   12     Number of cases listed:   12

-> COMMENT  In the next TWO lines, for a given data set put the actual
-> COMMENT  in place of the number '12' used for the example data set.

-> loop #i=1 to 12 .

-> COMMENT  In the next line, change '3' to whatever is the number
-> COMMENT  of variables.
-> COMMENT       The p critical value of chi square for a given case
-> COMMENT  is set as [the case number (after sorting) - .5] / the
-> COMMENT  sample size].

-> compute p=($casenum - .5) / 12. .

-> compute chisq=idf.chisq(p,3) .

-> end loop .

-> print formats p chisq (F8.5) .

-> list variables=y p mahal chisq/cases=9999/format=numbered .
          Y        P        MAHAL     CHISQ
   1      6    .04167      .70038    .30897
   2     11    .12500     1.65042    .69236
   3      5    .20833     1.98854   1.03962
   4      8    .29167     2.17303   1.38807
   5     12    .37500     2.19634   1.75398
   6      7    .45833     2.22174   2.15099
   7      4    .54167     2.37118   2.59519
   8      2    .62500     2.53196   3.10983
   9      1    .70833     2.59346   3.73392
  10     10    .79167     3.12622   4.54475
  11      9    .87500     5.59246   5.73941
  12      3    .95833     5.85428   8.22056

Number of cases read:   12     Number of cases listed:   12
```

(case # after sorting - .5) / n

```
-> plot
->   vertical='chi square'/
->   horizontal='Mahalabis distance'/
->   plot=chisq with mahal .
```

Hi-Res Chart # 6:Plot of chisq with mahal

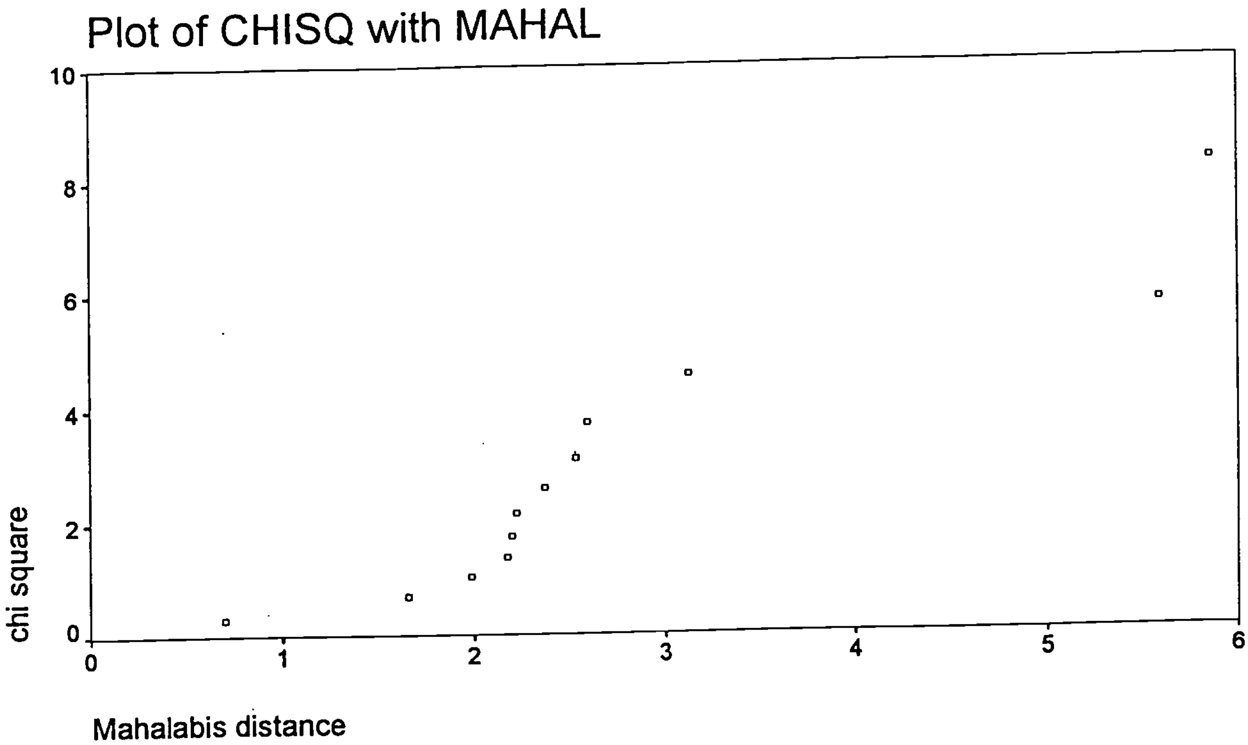

Plot of CHISQ with MAHAL
chi square
Mahalabis distance
0
2
4
6
8
10
0
1
2
3
4
5
6

Table 1

SPSS Descriptives Printout for a Variable with 100 Responses Demonstrating Normality

X

Valid cases: 100.0 Missing cases: .0 Percent missing: .0

Mean	.0000	Std Err	.1005	Min	-2.6000	Skewness	.0000
Median	.0000	Variance	1.0099	Max	2.6000	S E Skew	.2414
5% Trim	.0000	Std Dev	1.0049	Range	5.2000	Kurtosis	-.0900
95% CI for Mean (-.1994, .1994)				IQR	1.4000	S E Kurt	.4783

	Statistic	df	Significance
K-S (Lilliefors)	.0253	100	> .2000

Table 2

SPSS Descriptives Printout for a Variable with 26 Responses Failing to Demonstrate Normality

```
    ONE

Valid cases:        26.0  Missing cases:        .0  Percent missing:      .0

Mean          6.4038  Std Err       .4171  Min          2.9000  Skewness       .9959
Median        6.0500  Variance     4.5228  Max         12.5000  S E Skew       .4556
5% Trim       6.2791  Std Dev      2.1267  Range        9.6000  Kurtosis      1.6858
95% CI for Mean (5.5449, 7.2628)           IQR          2.8250  S E Kurt       .8865

                          Statistic        df        Significance

Shapiro-Wilks                 .9424        26             .2169
K-S (Lilliefors)              .1151        26           > .2000
```

Table 3

Vocabulary and Math Scores from 10 students

Pupil Number	Vocabulary Test (X_1)	Math Test (X_2)
1	19	15
2	20	18
3	17	18
4	16	12
5	19	16
6	17	16
7	18	13
8	17	20
9	15	17
10	18	16
Mean	17.6	16.1

Table 4

SPSS Descriptives Printout forVariables One, Two, and Three of Multinor data

			Statistic	Std. Error
ONE	Mean		6.4038	.4171
	95% Confidence Interval for Mean	Lower Bound	5.5449	
		Upper Bound	7.2628	
	5% Trimmed Mean		6.2791	
	Median		6.0500	
	Variance		4.523	
	Std. Deviation		2.1267	
	Minimum		2.90	
	Maximum		12.50	
	Range		9.60	
	Interquartile Range		2.8250	
	Skewness		.996	.456
	Kurtosis		1.686	.887
TWO	Mean		6.8692	.5339
	95% Confidence Interval for Mean	Lower Bound	5.7695	
		Upper Bound	7.9689	
	5% Trimmed Mean		6.8474	
	Median		7.1000	
	Variance		7.413	
	Std. Deviation		2.7226	
	Minimum		3.00	
	Maximum		11.20	
	Range		8.20	
	Interquartile Range		5.6750	
	Skewness		.069	.456
	Kurtosis		-1.380	.887
THREE	Mean		6.7154	.3568
	95% Confidence Interval for Mean	Lower Bound	5.9805	
		Upper Bound	7.4502	
	5% Trimmed Mean		6.6440	
	Median		6.5500	
	Variance		3.310	
	Std. Deviation		1.8194	
	Minimum		4.20	
	Maximum		11.00	
	Range		6.80	
	Interquartile Range		2.9750	
	Skewness		.344	.456
	Kurtosis		-.506	.887

Table 5

Tests of Normality for Variables One, Two, and Three

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ONE	.115	26	.200*	.942	26	.217
TWO	.122	26	.200*	.925	26	.069
THREE	.094	26	.200*	.950	26	.310

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

BEST COPY AVAILABLE

Figure Captions

Figure 1. Q-Q plot of 100 responses to a variable demonstrating normality.

Figure 2. Q-Q plots of 50 responses to a variable failing to demonstrate normality.

Figure 3. Stem and leaf plot of 100 responses to a variable demonstrating normality.

Figure 4. Histogram of 100 responses to a variable demonstrating normality.

Figure 5. Stem and leaf plots of 50 responses to a variable failing to demonstrate normality.

Figure 6. Histograms of 50 responses to a variable failing to demonstrate normality.

Figure 7. Scattergram of vocabulary and math scores.

Note. From Selected Topics in Advanced Statistics: An Elementary Approach (p.15), by M. Tatsuoka, 1971, Champaign, Illinois: The Institute for Personality and Ability Testing. Copyright 1971 by the Institute for Personality and Ability Testing.

Figure 8. Graphical representation of a bivariate normal distribution (1)

Note. From Selected Topics in Advanced Statistics: An Elementary Approach (p.16), by M. Tatsuoka, 1971, Champaign, Illinois: The Institute for Personality and Ability Testing. Copyright 1971 by the Institute for Personality and Ability Testing.

Figure 9. Graphical representation of a bivariate normal distribution (2)

Note. From Multivariate Analysis: Techniques for Educational Psychological Research (p. 64), by M. Tatsuoka, 1971, New York: John Wiley & Sons. Copyright 1971 by John Wiley & Sons Inc.

Figure 10. Graphical representation of a bivariate normal distribution (3)

Note. From Multivariate Statistical Methods: An Introduction (p. 52), by M. Karson, 1982, Ames, Iowa: The Iowa State University Press. Copyright 1982 by The Iowa State University

Press.

Figure 11. Graphical representation of a bivariate normal distribution (4)

Note. From Applied Linear Statistical Models (p. 633), by J. Neter, M. Kutner, C. Nachtsheim, and W. Wasserman, Chicago: Irwin. Copyright 1996 by Times Mirror Higher Education Group, Inc.

Figure 12. Contour diagram for a bivariate normal surface

Note. From Applied Linear Statistical Methods (p. 26), by D. Morrison, 1983, Englewood Cliffs, New Jersey: Prentice-Hall, Inc. Copyright 1983 by Prentice-Hall, Inc.

Figure 13. Q-Q plot of variable one of Multinor data

Figure 14. Q-Q plot of variable two of Multinor data

Figure 15. Q-Q plot of variable three of Multinor data

Figure 16. Stem and leaf plot of variable one of Mulitinor data

Figure 17. Stem and leaf plot of variable two of Mulitinor data

Figure 18. Stem and leaf plot of variable three of Mulitinor data

Figure 19. Histogram of variable one of Multinor data

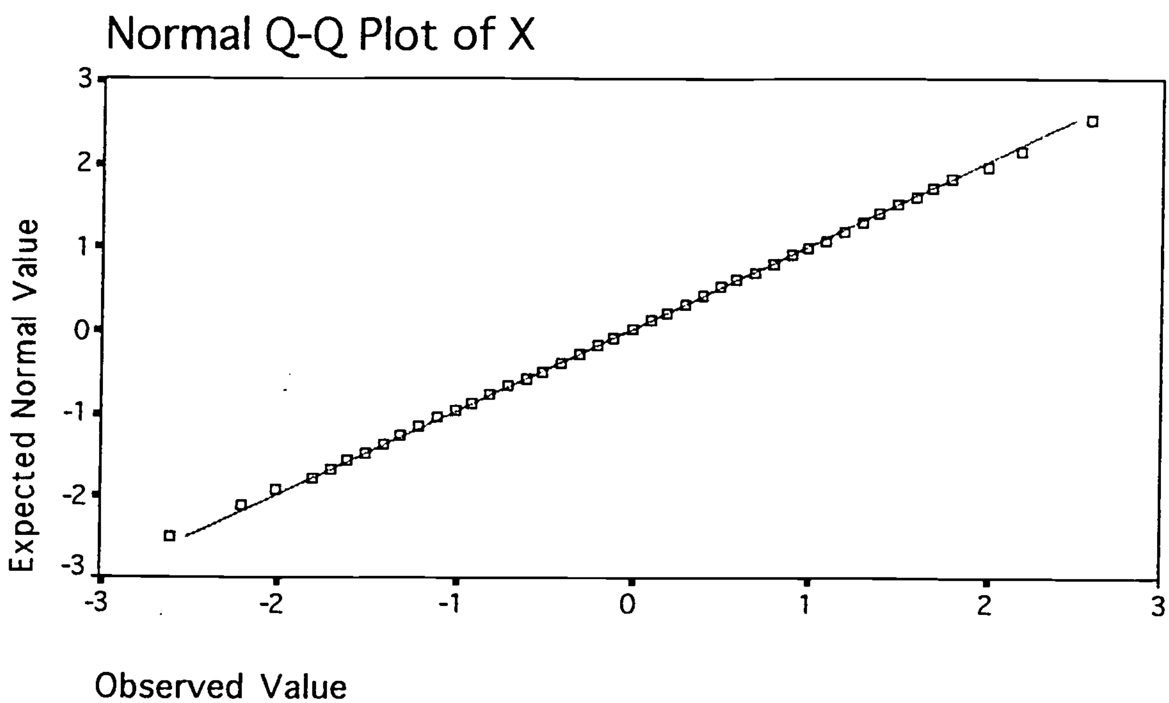
Figure 20. Histogram of variable two of Multinor data

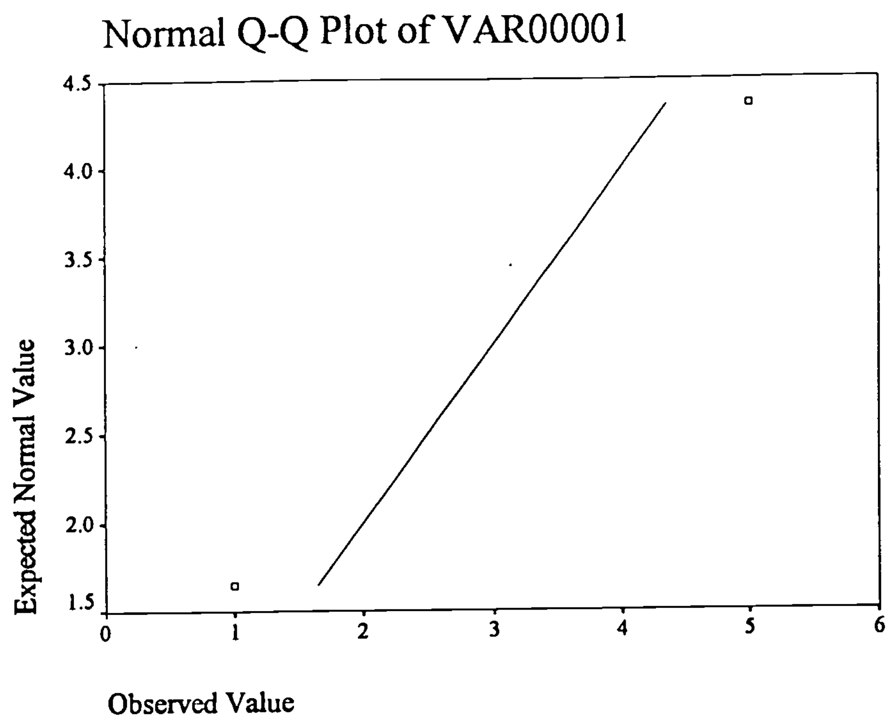
Figure 21. Histogram of variable three of Multinor data

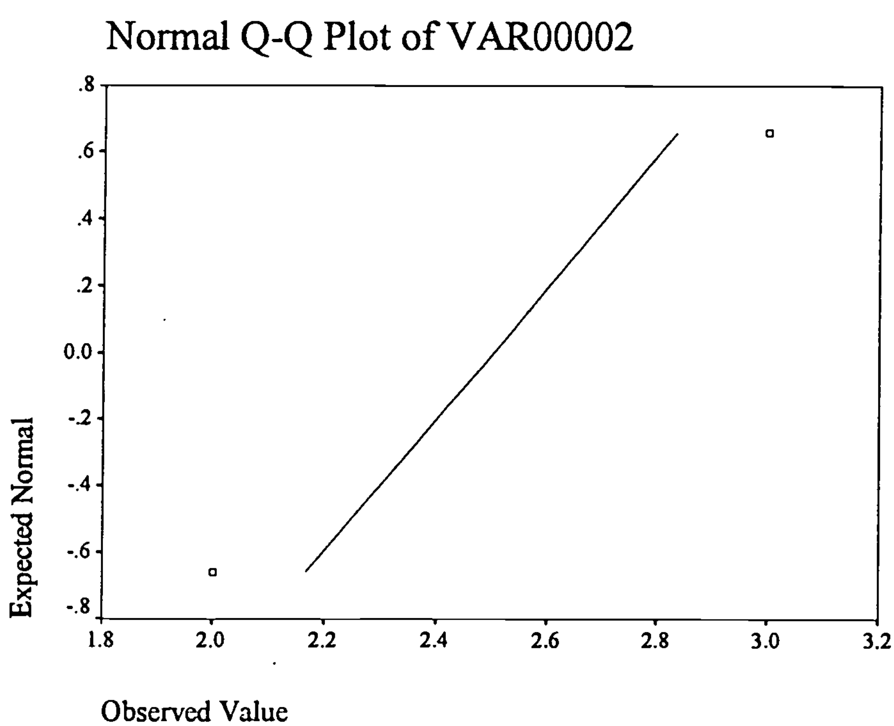
Figure 22. Scattergram of variables one and three of Multinor data

Figure 23. Scattergram of variables one and two of Multinor data

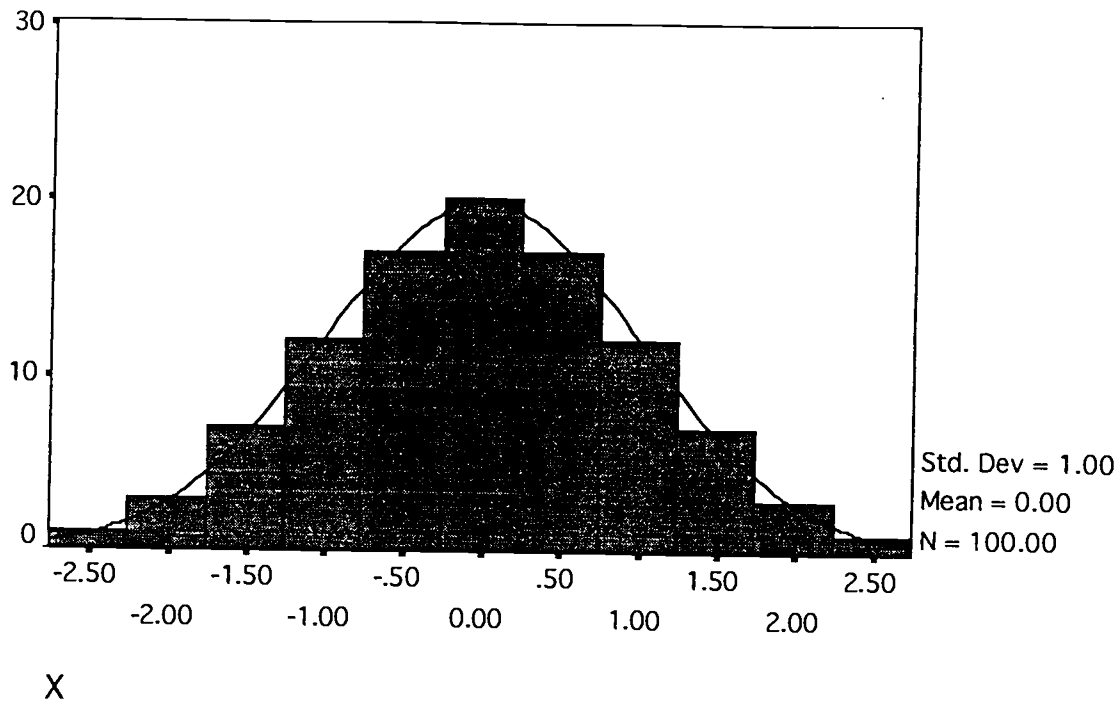
Figure 24. Scattergram of variables two and three of Multinor data


Normal Q-Q Plot of X
Expected Normal Value
Observed Value
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


Normal Q-Q Plot of VAR00001
Expected Normal Value
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0
1
2
3
4
5
6
Observed Value

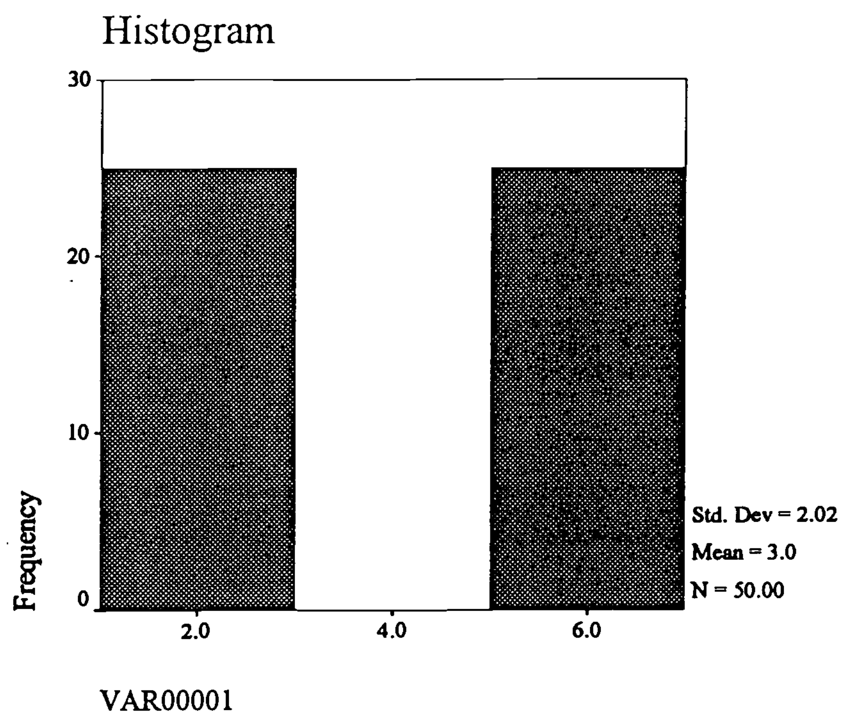

Normal Q-Q Plot of VAR00002
Expected Normal
.8
.6
.4
.2
0.0
-.2
-.4
-.6
-.8
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2
Observed Value

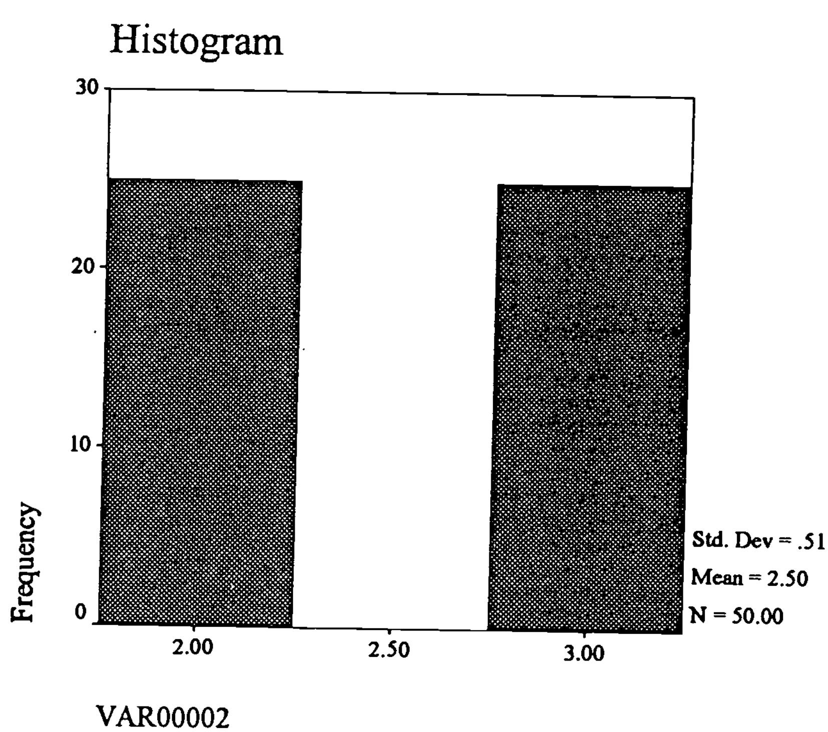
```
Frequency    Stem &  Leaf

    1.00       -2 .  6
    2.00       -2 *  02
    4.00       -1 .  5678
   10.00       -1 *  0011223344
   15.00       -0 .  555666777888999
   16.00       -0 *  1111222233334444
   20.00        0 *  00001111222233334444
   15.00        0 .  555666777888999
   10.00        1 *  0011223344
    4.00        1 .  5678
    2.00        2 *  02
    1.00        2 .  6

Stem width:      1.00
Each leaf:       1 case(s)
```

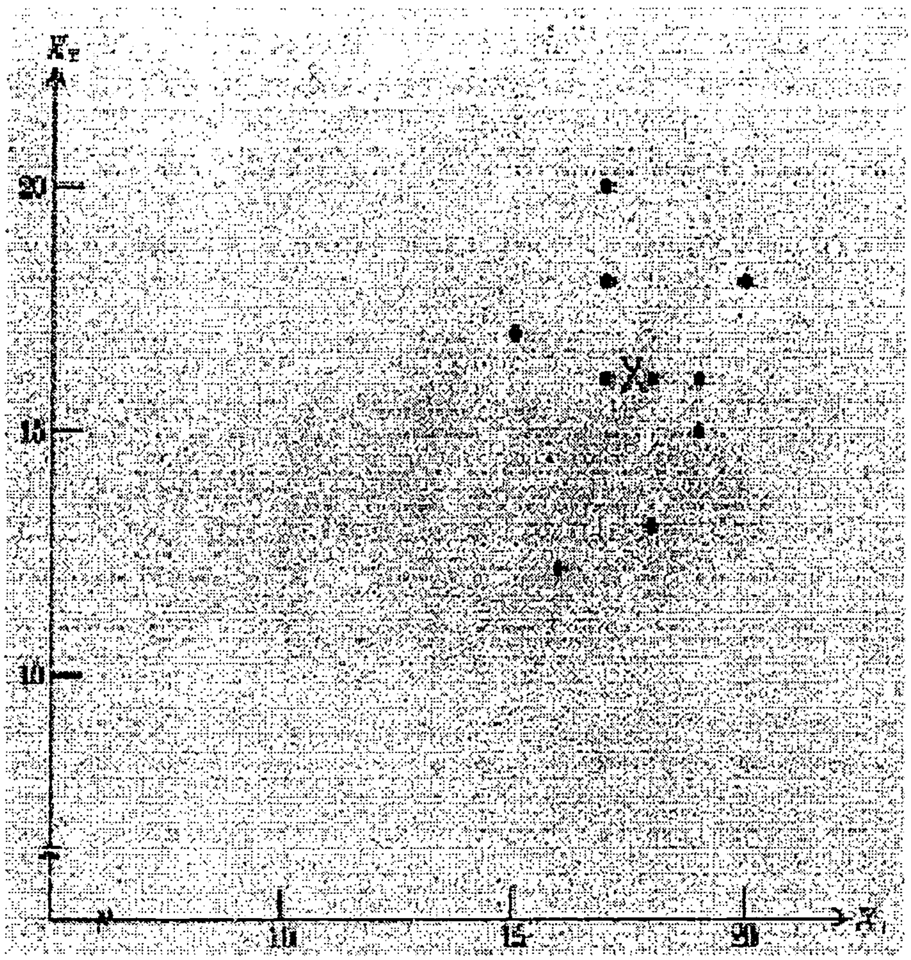

30
20
10
0
Std. Dev = 1.00
Mean = 0.00
N = 100.00
-2.50
-1.50
-.50
.50
1.50
2.50
-2.00
-1.00
0.00
1.00
2.00
X

```
VAR00001 Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

    25.00        1 .  0000000000000000000000000
      .00        1 .
      .00        2 .
      .00        2 .
      .00        3 .
      .00        3 .
      .00        4 .
      .00        4 .
    25.00        5 .  0000000000000000000000000

 Stem width:      1.00
 Each leaf:       1 case(s)
```

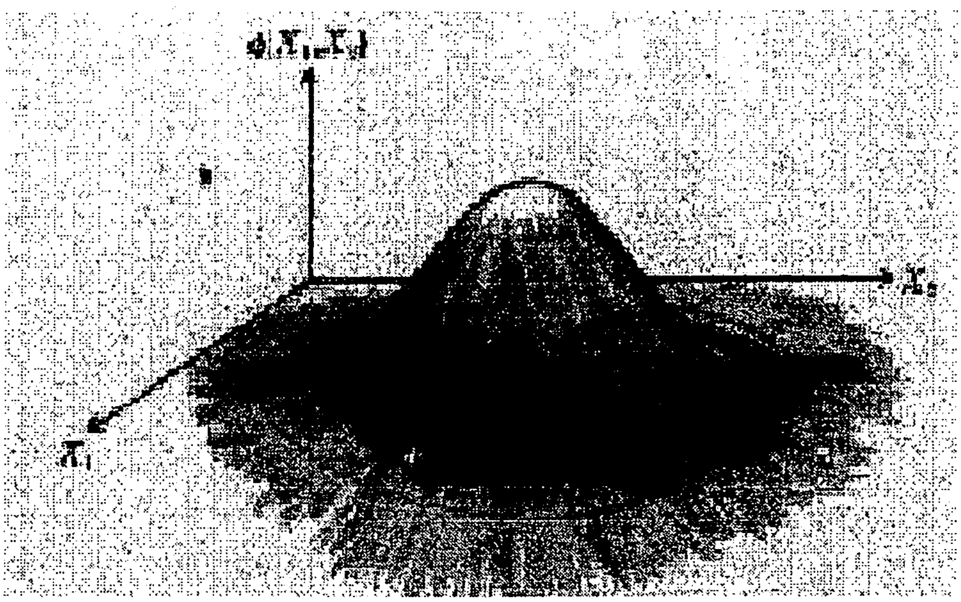
```
VAR00002 Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

    25.00        2 .  0000000000000000000000000
      .00        2 .
      .00        2 .
      .00        2 .
      .00        2 .
    25.00        3 .  0000000000000000000000000

 Stem width:      1.00
 Each leaf:       1 case(s)
```


Histogram
30
20
10
0
Frequency
Std. Dev = 2.02
Mean = 3.0
N = 50.00
2.0
4.0
6.0
VAR00001

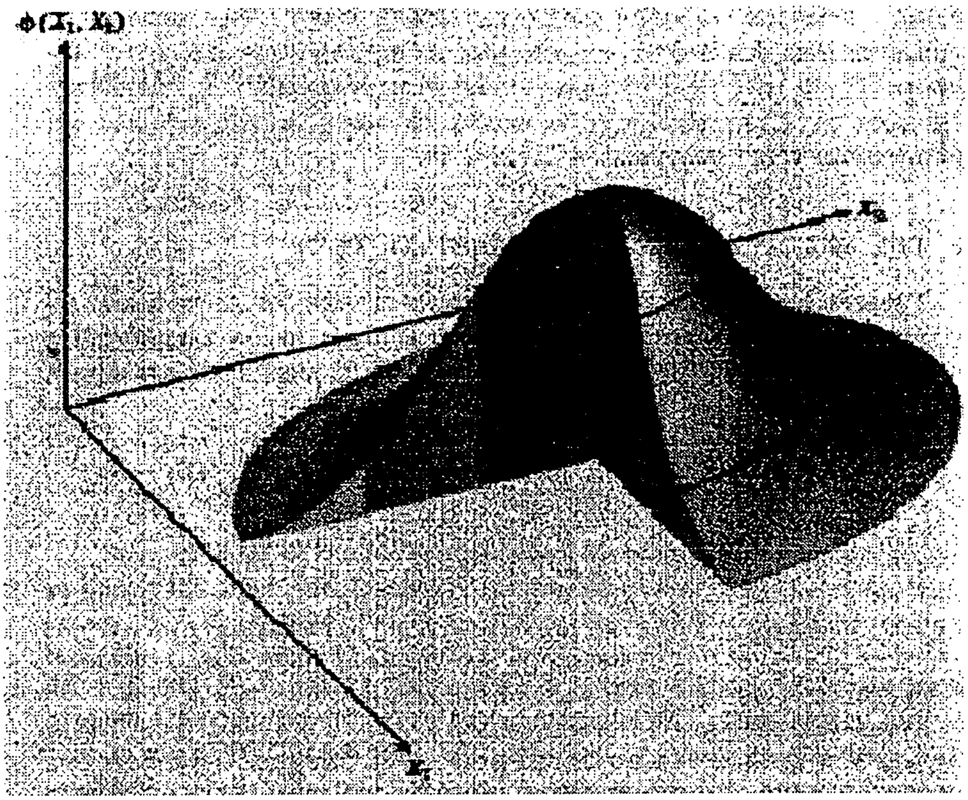

Histogram
30
20
10
0
Frequency
Std. Dev = .51
Mean = 2.50
N = 50.00
2.00
2.50
3.00
VAR00002

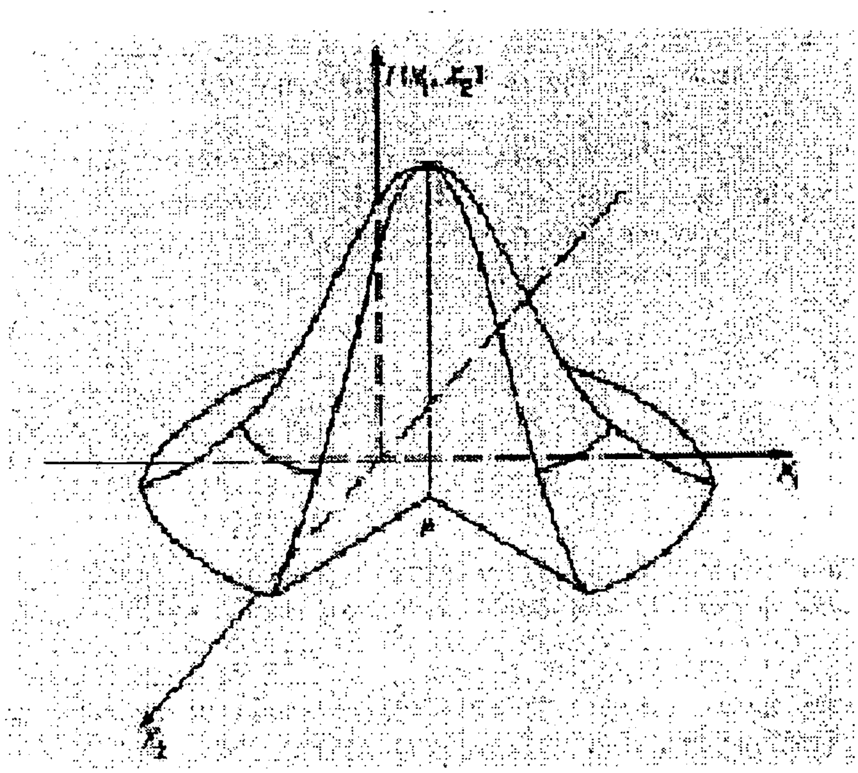


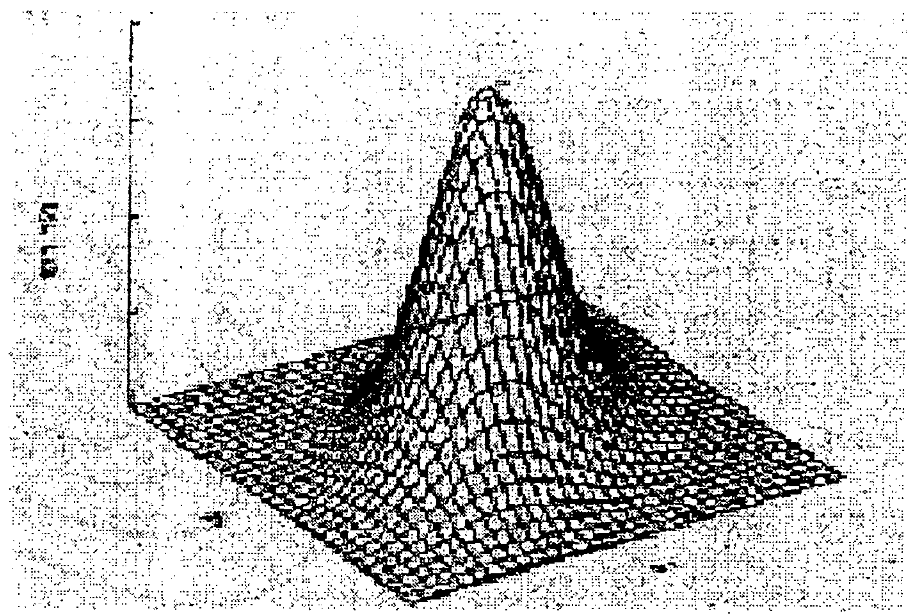
BEST COPY AVAILABLE

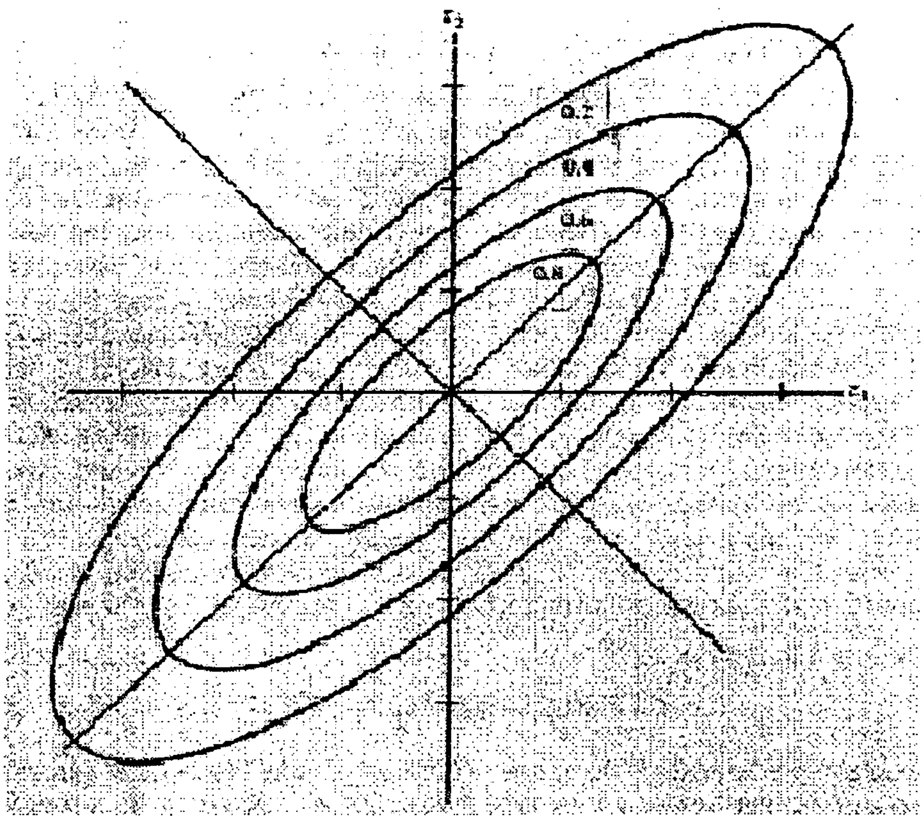


BEST COPY AVAILABLE

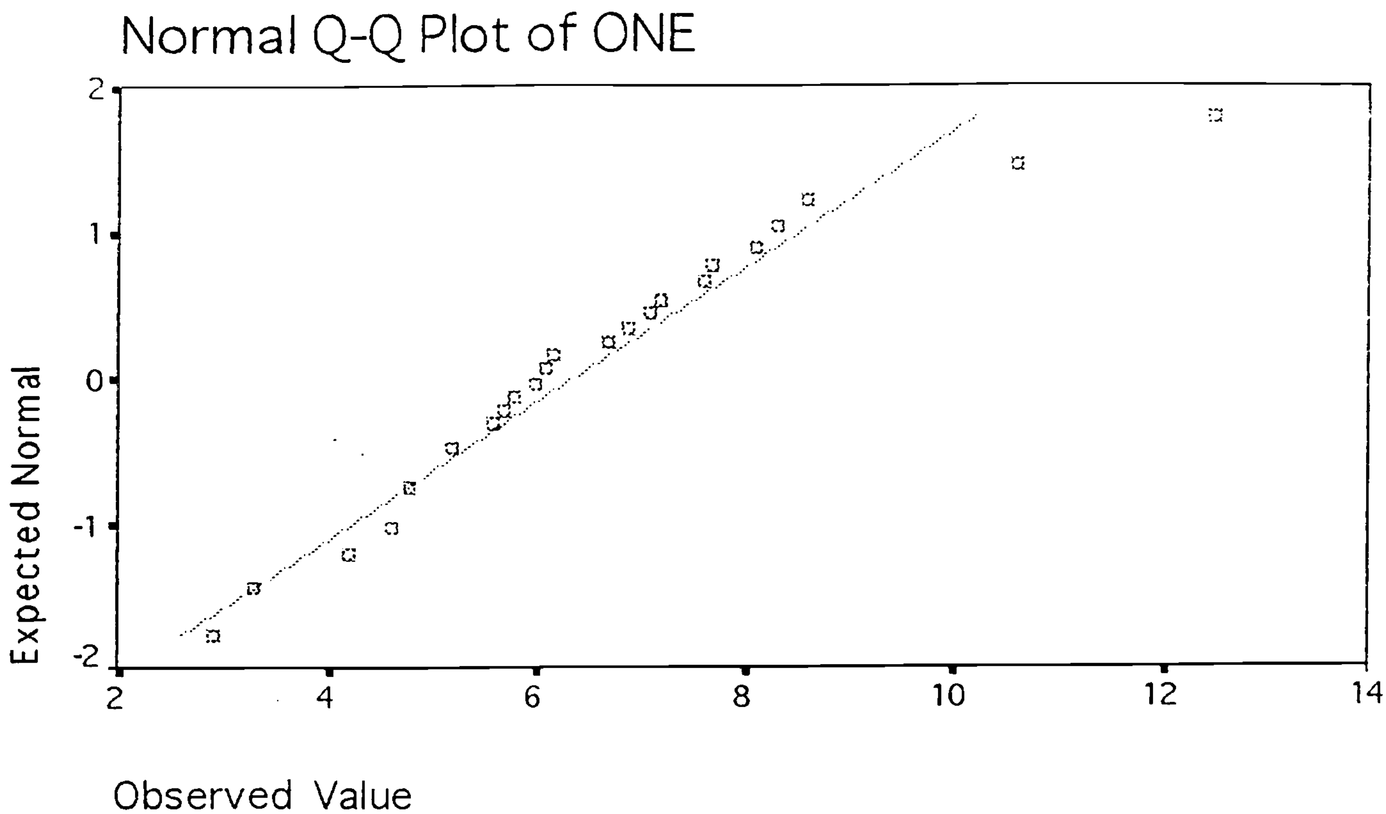

$\phi(X_1, X_2)$
X_2
X_1


$f(X_1, X_2)$
X_1
X_2

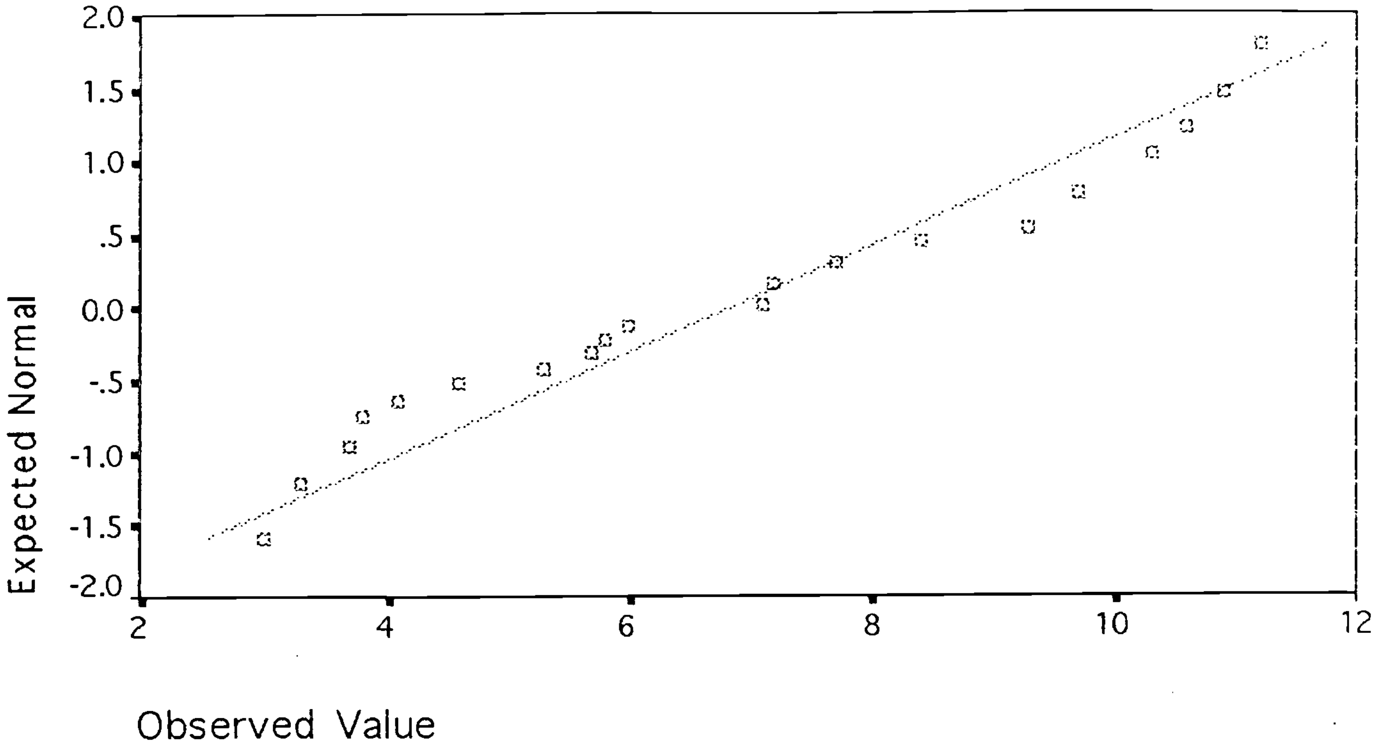


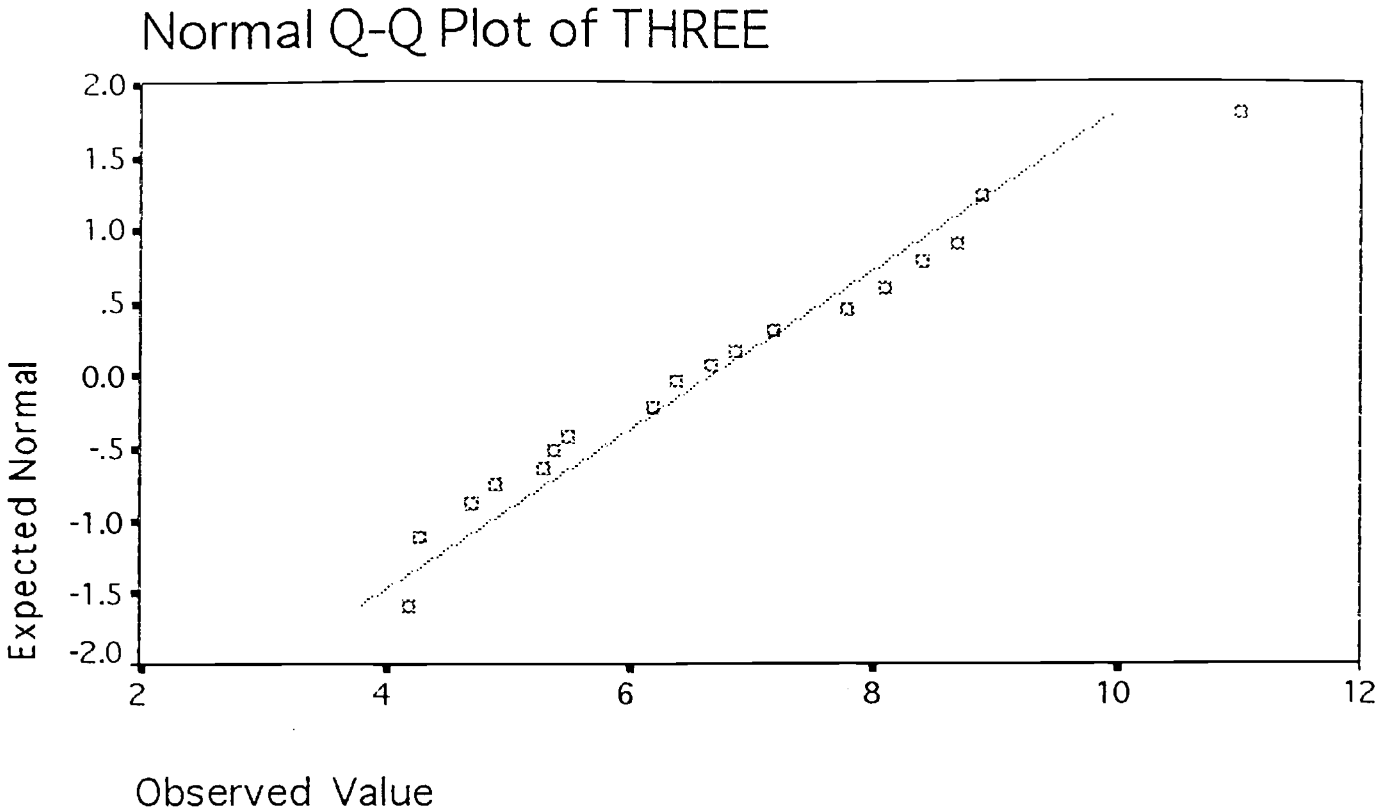


BEST COPY AVAILABLE


Normal Q-Q Plot of ONE
Expected Normal
2
1
0
-1
-2
2
4
6
8
10
12
14
Observed Value

Normal Q-Q Plot of TWO

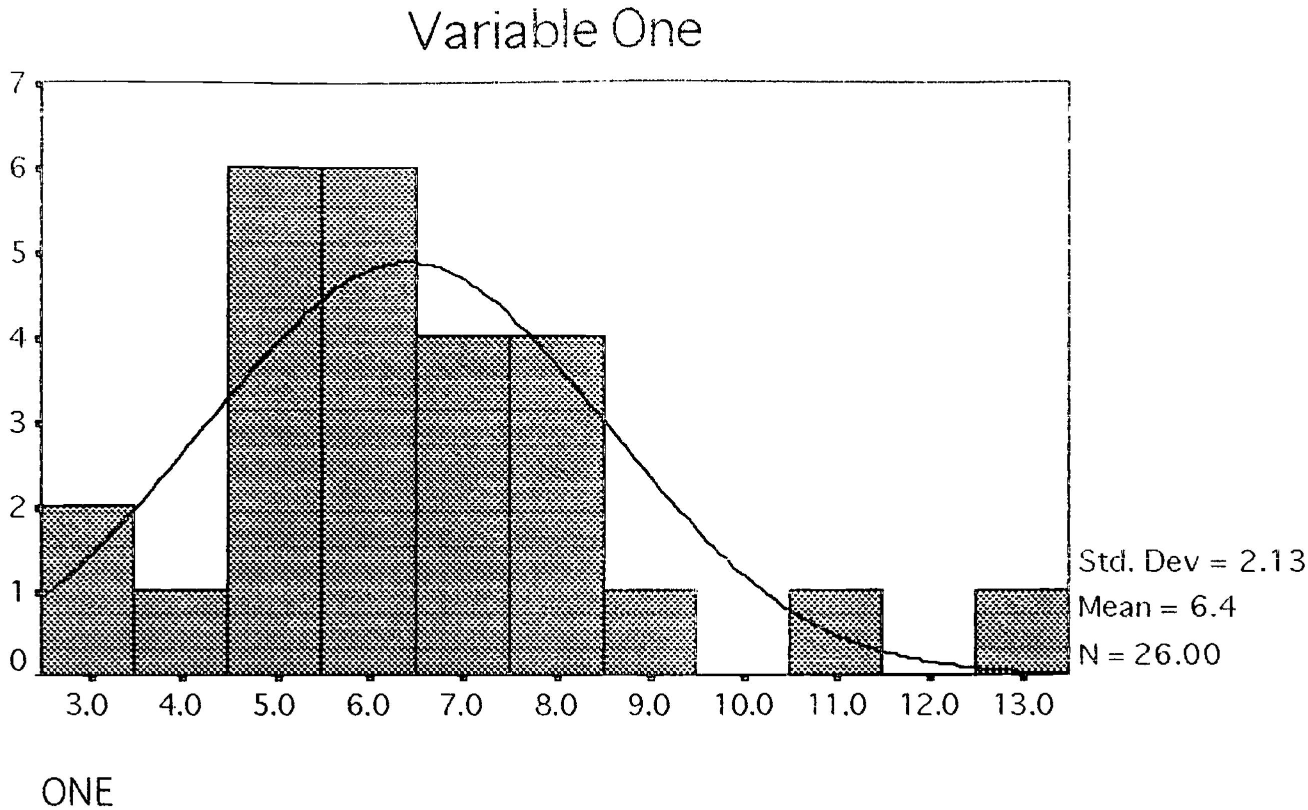



Normal Q-Q Plot of THREE
Expected Normal
2.0
1.5
1.0
.5
0.0
-.5
-1.0
-1.5
-2.0
2
4
6
8
10
12
Observed Value

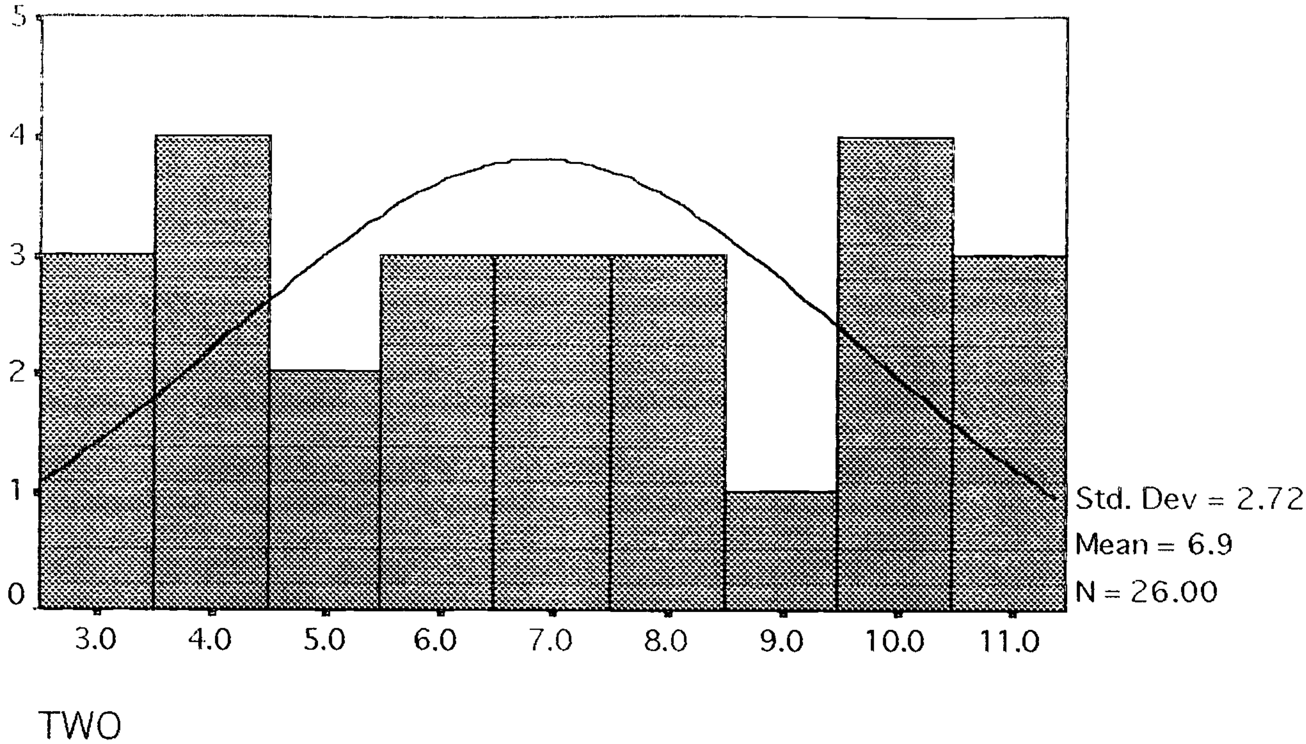
```
Frequency    Stem &  Leaf

    1.00        2 .  9
    1.00        3 .  3
    5.00        4 .  26888
    5.00        5 .  22678
    5.00        6 .  01279
    4.00        7 .  1267
    3.00        8 .  136
     .00        9 .
    1.00       10 .  6
    1.00 Extremes    (12.5)

Stem width:       1.00
Each leaf:        1 case(s)
```

```
Frequency    Stem &  Leaf

    6.00        3 .  003778
    2.00        4 .  16
    3.00        5 .  378
    1.00        6 .  0
    5.00        7 .  11277
    1.00        8 .  4
    4.00        9 .  3777
    3.00       10 .  369
    1.00       11 .  2

Stem width:       1.00
Each leaf:        1 case(s)
```

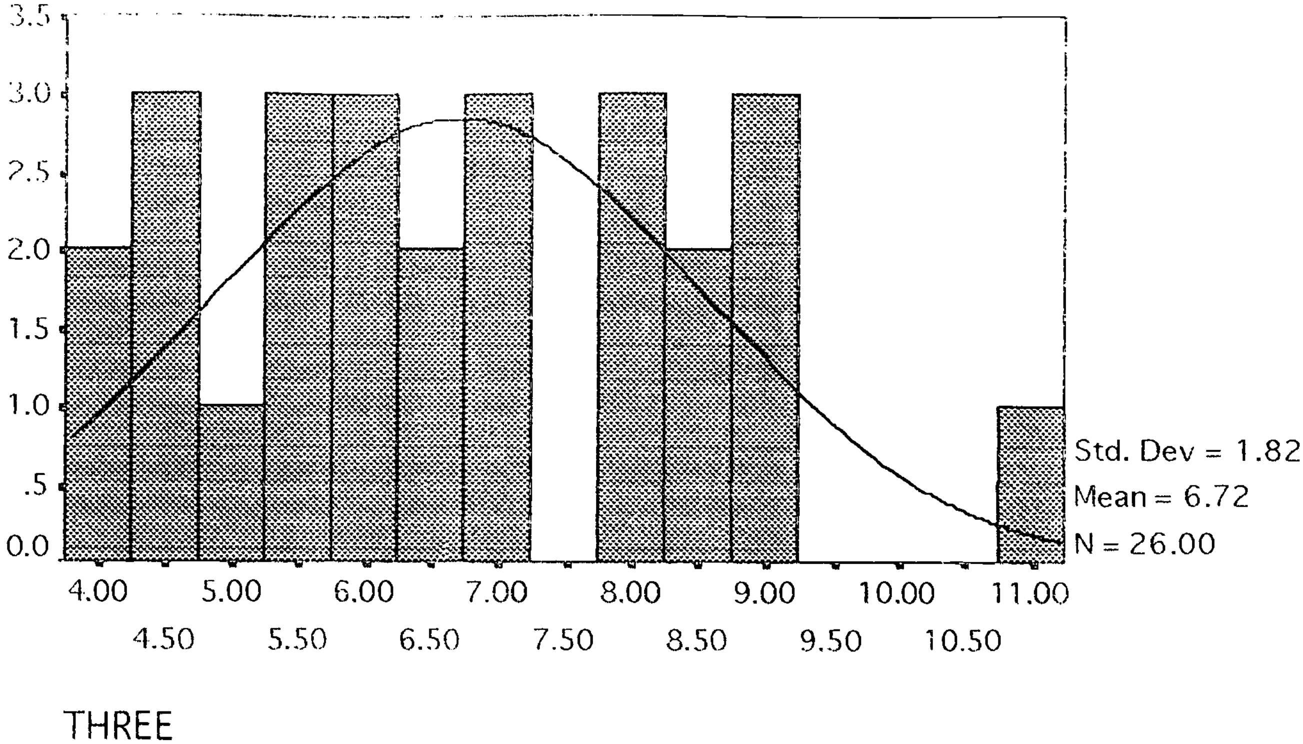
```
Frequency    Stem &  Leaf

    6.00        4 .  223379
    3.00        5 .  345
    6.00        6 .  222479
    3.00        7 .  228
    7.00        8 .  1147999
     .00        9 .
     .00       10 .
    1.00       11 .  0

Stem width:      1.00
Each leaf:       1 case(s)
```

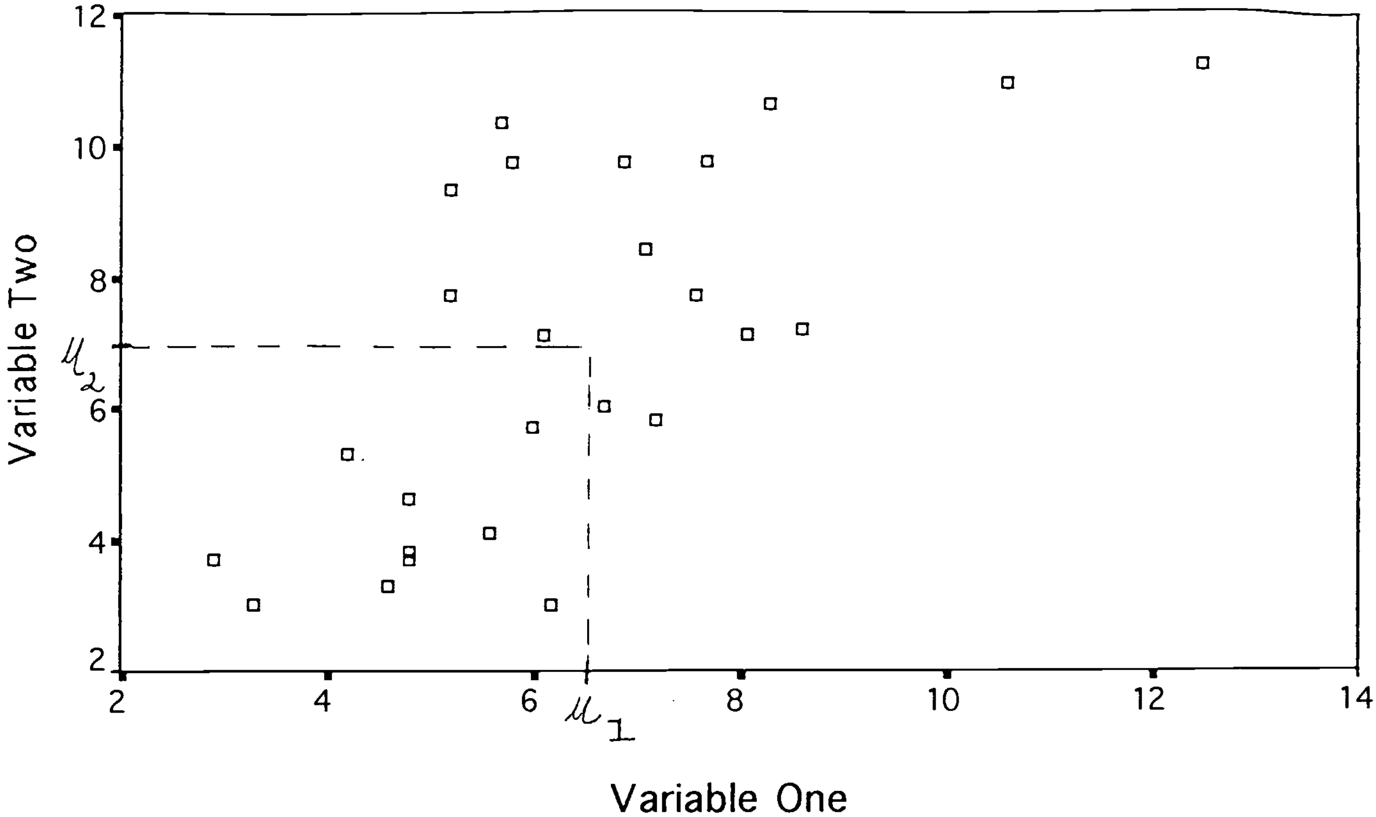

Variable One
7
6
5
4
3
2
1
0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
Std. Dev = 2.13
Mean = 6.4
N = 26.00
ONE

Variable Two

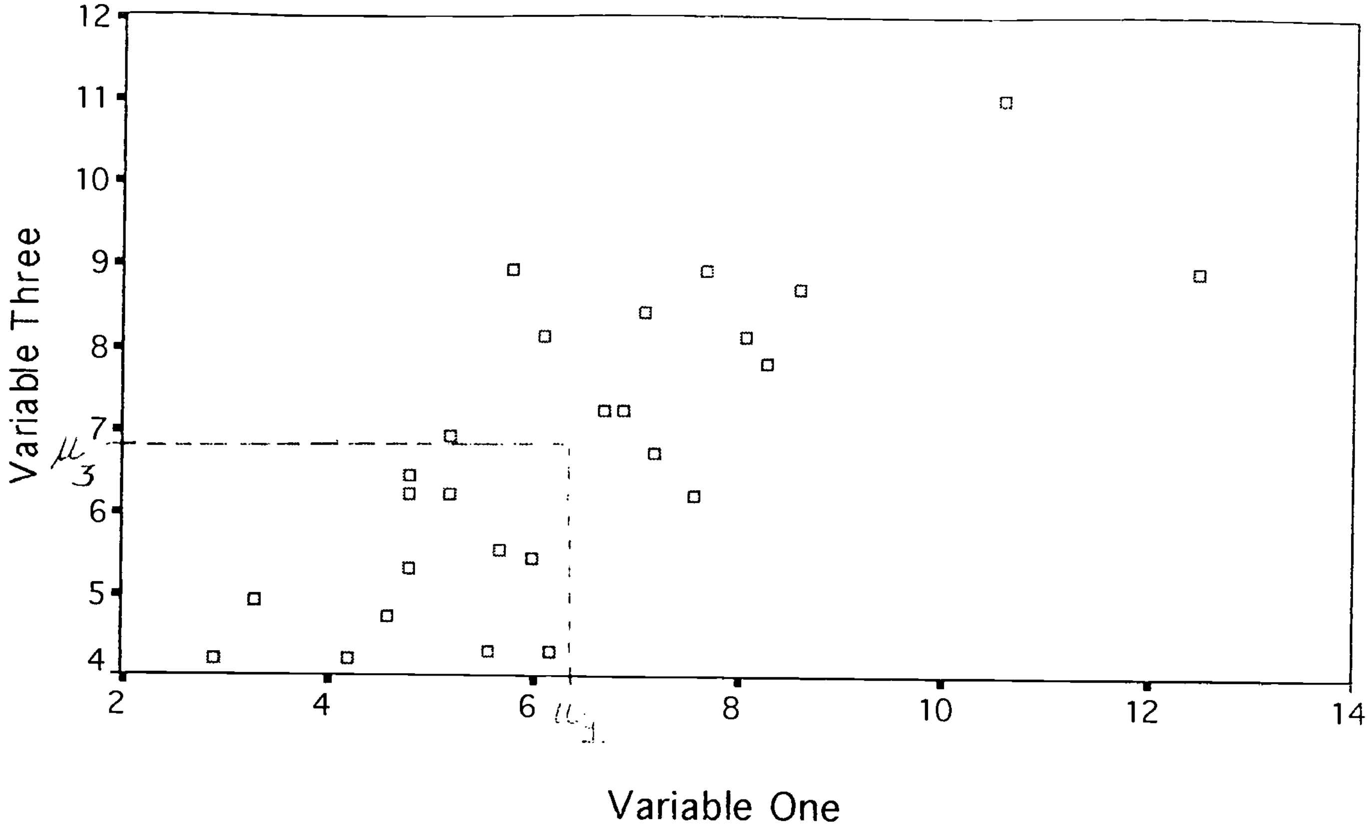


Variable Three

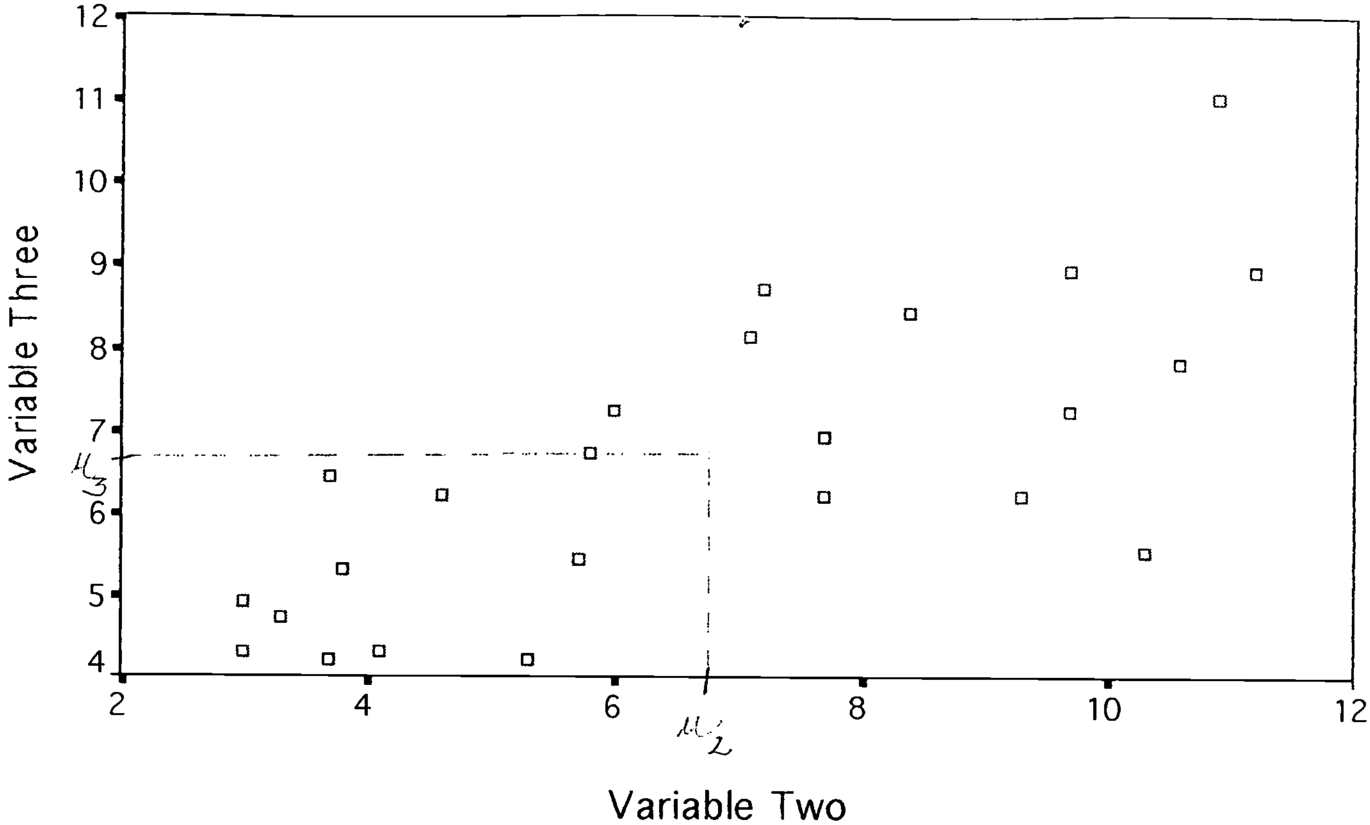




12
10
8
6
4
2
Variable Two
μ_2
2
4
6
μ_1
8
10
12
14
Variable One



Variable Three
Variable One
12
11
10
9
8
7
6
5
4
2
4
6
8
10
12
14
μ_3
μ_1


12
11
10
9
8
7
6
5
4
Variable Three
2
4
6
8
10
12
Variable Two

TM028188

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Ways to Evaluate the Assumption of Multivariate Normality	
Author(s): Alyce Smithson Ashcraft	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] ← Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
ALYCE S. ASHCRAFT
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document → [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here
Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X [signature]	Position: Doctoral Student
Printed Name: Alyce S. Ashcraft	Organization: University of Texas at Austin
Address: 1502 Richland Ct N, College Station, TX 77845	Telephone Number: (409) 690-0775
	Date: 2/23/98

OVER

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC , *or*, if you wish ERIC to cite the availability of this document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents which cannot be made available through EDRS).

Publisher/Distributor:	
Address:	
Price Per Copy:	Quantity Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name and address of current copyright/reproduction rights holder:

Name:

Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

If you are making an unsolicited contribution to ERIC, you may return this form (and the document being contributed) to:

ERIC Facility
1301 Piccard Drive, Suite 300
Rockville, Maryland 20850-4305
Telephone: (301) 258-5500